

TWO WEIGHT INEQUALITY FOR THE HILBERT TRANSFORM: A REAL VARIABLE CHARACTERIZATION

MICHAEL T. LACEY, ERIC T. SAWYER, CHUN-YEN SHEN, AND IGNACIO URIARTE-TUERO

ABSTRACT. Let σ and w be locally finite positive Borel measures on \mathbb{R} which do not share a common point mass. Then, the Hilbert transform $H(\sigma f)$ maps from $L^2(\sigma)$ to $L^2(w)$ if and only if $H(\sigma f)$ maps $L^2(\sigma)$ into weak- $L^2(w)$, and the dual weak-type inequality holds. This is a corollary to a more precise characterization in terms of a Poisson A_2 condition on the pair of weights, and conditions phrased in terms of testing the norm inequality over any subset of an arbitrary interval.

CONTENTS

1.	Introduction	1
2.	Dyadic Grids and Haar Functions	4
3.	The Parallel Corona	6
4.	A Function Theory	10
5.	Energy, Monotonicity, and Poisson	16
6.	Testing Minimal Bounded Fluctuation Functions	18
7.	The Functional Energy Inequality	24
	References	33

1. INTRODUCTION

Let $H\nu(x) = \text{p.v.} \int_{\mathbb{R}} \frac{d\nu(y)}{y-x}$ be the Hilbert transform of the measure ν . In this definition, the principal value need not exist, so we always understand that there is some standard truncation of the integral in place, and all relevant estimates are assumed to be independent of how the truncation is taken. Given weights (i.e. locally bounded positive Borel measures) σ and w on the real line \mathbb{R} , we consider the following *two weight norm inequality* for the Hilbert transform,

$$(1.1) \quad \int_{\mathbb{R}} |H(f\sigma)|^2 dw \leq N^2 \int_{\mathbb{R}} |f|^2 d\sigma, \quad f \in L^2(\sigma),$$

M.L. supported in part by the NSF grant 0968499, and a grant from the Simons Foundation (#229596 to Michael Lacey).

Research supported in part by NSERC.

Research supported in part by the NSF, through grant DMS-0901524.

where \mathcal{N} is the best constant in the inequality, uniform over all truncations of the Hilbert transform kernel. A question due to Nazarov-Treil-Volberg, see [29], is whether or not (1.1) is equivalent to the following necessary conditions. The (half-Poisson) A_2 condition

$$P(\sigma, I) \frac{w(I)}{|I|} \leq \mathcal{A}_2, \quad \frac{\sigma(I)}{|I|} P(w, I) \leq \mathcal{A}_2,$$

where the inequalities are uniform over intervals I , and $P(\sigma, I)$ is the usual Poisson extension of σ evaluated at point in the upper half-plane $(x_I, |I|)$, where x_I is the center of I . And the interval testing conditions

$$\int_I |H(\mathbf{1}_I \sigma)|^2 dw \leq \mathcal{T}^2 \sigma(I), \quad \int_I |H(\mathbf{1}_I w)|^2 d\sigma \leq \mathcal{T}^2 w(I),$$

also holding uniformly over all intervals I . The norm inequality above is phrased in a self-dual fashion, namely the dual inequality is obtained by interchanging the weights w and σ . Thus, the two testing conditions above are dual. The best constants in the two inequalities can be of different orders of magnitude.

In this paper we prove a weaker variant of this conjecture, with the two interval testing conditions replaced with stronger testing conditions on bounded functions.

Theorem 1.3. *Let σ and w be locally finite positive Borel measures on the real line \mathbb{R} with no common point masses. There holds*

$$\mathcal{N} \approx \mathcal{A}_2^{1/2} + \mathcal{T}_\infty,$$

where the latter constant is the best constant in the inequalities, below.

$$\begin{aligned} \int_I H(\sigma f \mathbf{1}_I)^2 dw &\leq \mathcal{T}_\infty^2 \|f\|_\infty^2 \sigma(I), \\ \int_I H(w g \mathbf{1}_I)^2 d\sigma &\leq \mathcal{T}_\infty^2 \|g\|_\infty^2 w(I). \end{aligned}$$

These hold uniformly over all intervals I , and functions f, g . Note that only the L^∞ norms of the functions enters into the right hand side.

A characterization in terms of weak-type norms follows.

Corollary 1.4. *Under the hypotheses above, there holds $\mathcal{N} \simeq \mathcal{W}$, where the latter constant is the best constant in the weak-type inequalities*

$$\|H(\sigma f)\|_{L^{2,\infty}(w)} \leq \mathcal{W} \|f\|_{L^2(\sigma)}, \quad \|H(w g)\|_{L^{2,\infty}(\sigma)} \leq \mathcal{W} \|g\|_{L^2(w)}.$$

That $\mathcal{N} \simeq \mathcal{A}_2^{1/2} + \mathcal{W}$ is the immediate corollary, using well-known duality properties of Lorentz spaces. But in addition, the half-Poisson constant satisfies $\mathcal{A}_2^{1/2} \lesssim \mathcal{W}$, as follows from inspection of the proof of $\mathcal{A}_2^{1/2} \lesssim \mathcal{N}$ in [8, Section 2].

The form of our Theorem and Corollary closely match results about positive operators [27], and the context of singular integrals, the main results of [21]. The latter paper focuses on the special case of $\sigma = 1/w$ with w an A_2 weight, and the Hilbert transform is replaced by an arbitrary

Calderón-Zygmund operator in any dimension, providing a sharp estimate of the norm of the operator in terms of the A_2 constant of w and the testing constant. This result was employed by Hytönen, in his resolution of the A_2 conjecture [4]. (Simpler proofs have subsequently been found.) Our main theorem bears strong resemblance to the characterizations of two weight inequalities found in [7].

We remark that our constant \mathcal{T}_∞ is comparable to the best constant in the inequalities

$$\int_I H(\sigma \mathbf{1}_F)^2 d\omega \leq \mathcal{T}_1^2 \sigma(I), \quad \int_I H(w \mathbf{1}_G)^2 d\sigma \leq \mathcal{T}_1^2 w(I),$$

where $F, G \subset I$, and I is an arbitrary interval.

In the circumstances in which the two weight problem arises, one would like sufficient conditions for the L^2 norm inequality that are as simple as possible, namely the interval testing condition above. Verifying that one has $\mathcal{T}_\infty \lesssim \mathcal{A}_2^{1/2} + \mathcal{T} := \mathcal{H}$ appears to require techniques beyond the scope of this paper. Also, certain complex variable characterizations of the two weight inequality were found by Cotlar-Sadosky, [1].

The Nazarov-Treil-Volberg conjecture has only been verified before under additional hypotheses on the pair of weights, hypotheses which are not necessary for the two weight inequality. The so-called pivotal condition of [29] is not necessary, as was proved in [8]. The pivotal condition is still an interesting condition: It is all that is needed to characterize the boundedness of the Hilbert transform, together with the Maximal Function in both directions. But, the boundedness of this triple of operators is decoupled in the two weight setting [24].

Our argument has these attributes.

- (1) Certain degeneracies of the pair of weights must be addressed, the contribution of the innovative 2004 paper of Nazarov-Treil-Volberg [17], also see [29], which was further sharpened with the property of *energy* in [8]. This theme is further developed herein.
- (2) Properties of the Hilbert transform must be carefully exploited. This was a key contribution of [8], and it is continued here. What was known before is listed in §5. This paper adds two additional properties to the list.
- (3) The proof should proceed through the analysis of the bilinear form $\langle H(\sigma f), g w \rangle$, as one expects certain paraproducts to appear. Still, the paraproducts have no canonical form, suggesting that the proof be highly non-linear in f and g . The non-linear point of view was initiated in [9], and is central to this paper. A particular feature of our arguments is a repeated appeal to certain *quasi-orthogonality* arguments, providing (many) simplifications over prior arguments. For instance, we never find ourselves constructing auxiliary measures, and verifying that they are Carleson, a frequent step in many related arguments.
- (4) Corona decompositions should be recursive. We herein establish the first such decomposition, called the *parallel corona*, see Theorem 3.7. The proof of this Theorem depends in a critical way on a new property of the Hilbert transform, the *functional energy inequality* of Theorem 7.4. Both of these are of independent interest.
- (5) There is a function theory relevant to non-doubling measure spaces in one dimension. An essential intermediate step is to provide (very sharp) sufficient conditions for the two

weight inequality in term of testing bounded functions, and functions of *minimal bounded fluctuation*, see Definition 4.5, and Theorem 4.8.

- (6) The testing constant for minimal bounded fluctuation functions is then shown to be dominated by the A_2 and the L^∞ testing constant, an argument that depends a remarkable property of minimal bounded fluctuation functions: Their martingale averages have finite variation.

One can phrase a two weight inequality question for any operator T , a question that became apparent with the foundational paper of Muckenhoupt [11] on A_p weights for the Maximal Function. Indeed, the case of Hardy's inequality was quickly resolved by Muckenhoupt [12]. The Maximal Function was resolved by one of us [26], and the fractional integrals, and, essential for this paper, Poisson integrals [27]. The latter paper established a result which closely paralleled the contemporaneous T1 theorem of David and Journé [2]. This connection, fundamental in nature, was not fully appreciated until the innovative work of Nazarov-Treil-Volberg [14–16] in developing a non-homogeneous theory of singular integrals. The two weight problem for dyadic singular integrals was only resolved recently [18]. Partial information about the two weight problem for singular integrals [21] was basic to the resolution of the A_2 conjecture [4], and several related results [5, 6, 21, 22]. Our result is the first real variable characterization of a two weight inequality for a continuous singular integral.

Interest in the two weight problem for the Hilbert transform arises from its natural occurrence in questions related to operator theory [20, 25], spectral theory [20], and model spaces [23], and analytic function spaces [10]. In the context of operator theory Sarason posed the conjecture (See [3].) that the Hilbert transform would be bounded if the pair of weights satisfied the (full) Poisson A_2 condition. This was disproved by Nazarov [13]. Advances on these questions have been linked to finer understanding of the two weight question, see for instance [19, 20], which build upon Nazarov's counterexample.

§2 introduces terminology associated with dyadic grids, and the basic notion of the good and bad intervals. Following that, the parallel corona is described. The function theory takes up §4, and this section concludes with the proof of Theorem 4.8, modulo the parallel corona. The proof of the latter depends critically on the functional energy inequality proved in §7. Estimating the minimal bounded fluctuation testing constant is in §6.

Acknowledgment. The authors benefited from a stimulating conference on two weight inequalities at the American Institute of Mathematics, Palo Alto California, in October 2011.

2. DYADIC GRIDS AND HAAR FUNCTIONS

2.1. Dyadic Grids. A collection of intervals \mathcal{G} is a *grid* if for all $G, G' \in \mathcal{G}$, we have $G \cap G' \in \{\emptyset, G, G'\}$. By a *dyadic grid* we mean a grid \mathcal{D} of intervals of \mathbb{R} such that for each interval $I \in \mathcal{D}$, the subcollection $\{I' \in \mathcal{D} : |I'| = |I|\}$ partitions \mathbb{R} , aside from endpoints of the intervals. In addition, the left and right halves of I , denoted by I_\pm , are also in \mathcal{D} .

For $I \in \mathcal{D}$, the left and right halves I_\pm are referred to as the *children* of I . We denote by $\pi_{\mathcal{D}}(I)$ the unique interval in \mathcal{D} having I as a child, and we refer to $\pi_{\mathcal{D}}I$ as the \mathcal{D} -parent of I .

We will work with subsets $F \subset \mathcal{D}$. We say that I has \mathcal{F} parent $\pi_{\mathcal{F}} I = F$ if $F \in \mathcal{F}$ is the minimal element of \mathcal{F} that contains I . The \mathcal{F} children of $F \in \mathcal{F}$ are the maximal $F' \in \mathcal{F}$ which are strictly contained in F .

2.2. Haar Functions. Let σ be a weight on \mathbb{R} , one that does not assign positive mass to any endpoint of a dyadic grid \mathcal{D} . If $I \in \mathcal{D}$ is such that σ assigns non-zero weight to both children of I , the associated Haar function is

$$h_I^\sigma := \sqrt{\frac{\sigma(I_-)\sigma(I_+)}{\sigma(I)}} \left(-\frac{I_-}{\sigma(I_-)} + \frac{I_+}{\sigma(I_+)} \right).$$

In this definition, we are identifying an interval with its indicator function, and we will do so throughout the remainder of the paper. This is an $L^2(\sigma)$ -normalized function, and has σ -integral zero. For any dyadic interval I_0 , it holds that $\{\sigma(I_0)^{-1/2} I_0\} \cup \{h_I^\sigma : I \in \mathcal{D}, I \subset I_0\}$ is an orthogonal basis for $L^2(I_0, \sigma)$.

We will use the notations $\hat{f}(I) = \langle f, h_I^\sigma \rangle_\sigma$, as well as

$$\Delta_I^\sigma f = \langle f, h_I^\sigma \rangle_\sigma h_I^\sigma = I_+ \mathbb{E}_{I_+}^\sigma f + I_- \mathbb{E}_{I_-}^\sigma f - I \mathbb{E}_I^\sigma f.$$

The second equality is the familiar martingale difference equality, and so we will refer to $\Delta_I^\sigma f$ as a martingale difference. It implies the familiar telescoping identity $\mathbb{E}_J^\sigma f = \sum_{I: I \supset J} \mathbb{E}_I^\sigma \Delta_I^\sigma f$.

For any function the *Haar support* of f is the collection $\{I \in \mathcal{D} : \hat{f}(I) \neq 0\}$.

2.3. Good-Bad Decomposition. With a choice of dyadic grid \mathcal{D} understood, we say that $J \in \mathcal{D}$ is (ϵ, r) -good if and only if for all intervals $I \in \mathcal{D}$ with $|I| \geq 2^{r+1}|J|$, the distance from J to the boundary of *either child* of I is at least $|J|^\epsilon |I|^{1-\epsilon}$.

For $f \in L^2(\sigma)$ we set $P_{\text{good}}^\sigma f = \sum_{\substack{I \in \mathcal{D} \\ I \text{ is } (\epsilon, r)\text{-good}}} \Delta_I^\sigma f$. The projection $P_{\text{good}}^w g$ is defined similarly.

To make the two reductions below, one must make a *random* selection of grids, as is detailed in [8, 29]. The use of random dyadic grids has been a basic tool since the foundational work of [14–16]. Important elements of the suppressed construction of random grids are that

- (1) It suffices to consider a single dyadic grid \mathcal{D} , but we will sometimes write \mathcal{D}^σ and \mathcal{D}^w to emphasize the role of the two weights.
- (2) For any fixed $0 < \epsilon < \frac{1}{2}$, we can choose integer r sufficiently large so that it suffices to consider f such that $\hat{f} = P_{\text{good}}^\sigma \hat{f}$, and likewise for $g \in L^2(w)$. Namely, it suffices to estimate the constant below, for arbitrary dyadic grid \mathcal{D} ,

$$|\langle H_\sigma f, g \rangle_w| \leq N_{\text{good}} \|f\|_\sigma \|g\|_w,$$

where it is required that $f = P_{\text{good}}^\sigma f \in L^2(\sigma)$ and $g = P_{\text{good}}^w g \in L^2(w)$.

That the functions are good is, at some moments, an essential property. We suppress it in notation, however taking care to emphasize in the text those places in which we appeal to the property of being good.

3. THE PARALLEL CORONA

With the notations of the previous section, the two weight inequality (1.1) is equivalent to boundedness of the bilinear form on $L^2(\sigma) \times L^2(w)$,

$$B(f, g) \equiv \langle H(f\sigma), g \rangle_w = \sum_{I \in \mathcal{D}^\sigma} \sum_{J \in \mathcal{D}^w} \langle H_\sigma(\Delta_I^\sigma f), \Delta_J^w g \rangle_w.$$

Here, and for the remainder of the paper, we use the notation $H_\sigma f = H(\sigma f)$. Also, as L^2 -norms predominate, we set $\|\phi\|_v := \|\phi\|_{L^2(v)}$.

Definition 3.1. We say that a dyadic interval $I \in \mathcal{D}^\sigma$ is *balanced* if

$$1/4 \leq \frac{-\inf_{x \in I} h_I^\sigma}{\sup_{x \in I} h_I^\sigma} \leq 4$$

Note that since h_I^σ has σ -mean zero, the infimum is necessarily negative. If I is not balanced, it is said to be *unbalanced*. An unbalanced interval I has children I_{small} and I_{large} satisfying $4\sigma(I_{\text{small}}) < \sigma(I_{\text{large}})$.

We say that $f \in L^2(\sigma)$ is *balanced* if the Haar support of \hat{f} only consists of balanced intervals. The definition of f being *unbalanced* is at slight variance: We say that f is *unbalanced* if for all intervals I in the Haar support of f we have

$$\sup_{x \in I} \Delta_I^\sigma f < -4 \inf_{x \in I} \Delta_I^\sigma f.$$

That is, the ‘large value’ of the martingale difference is negative. It is clear that any $f \in L^2(\sigma)$ is the orthogonal sum of $f_b + f_{u,1} - f_{u,2}$, where f_b is balanced, and $f_{u,j}$ are unbalanced. It suffices to assume that f is either balanced, or unbalanced, and likewise for g .

A particular feature of this paper is a careful analysis of the unbalanced functions. This terminology will help the reader identify these portions of the proofs below.

Definition 3.2. A collection \mathcal{F} of dyadic intervals is σ -Carleson if

$$(3.3) \quad \sum_{F \in \mathcal{F}: F \subset S} \sigma(F) \leq C_{\mathcal{F}} \sigma(S), \quad S \in \mathcal{F}.$$

The constant $C_{\mathcal{F}}$ is referred to as the Carleson norm of \mathcal{F} .

Throughout, we can take $C_{\mathcal{F}}$ to be a fixed constant. We will work with two functions that are supported on an interval I_0 , that will change repeatedly. Let $L_0^2(I_0, \sigma)$ be functions in $L^2(\sigma)$ supported on I_0 and have $\int_{I_0} f \, d\sigma = 0$. It is very easy to reduce to the case of f and g being of integral zero in their respective spaces, and so we always assume this.

Definition 3.4. For fixed constants $C_{CZ} \geq 4$, we call $\mathcal{F} \subset \mathcal{D}^\sigma$ and non-negative numbers $\{\alpha_f(F) : F \in \mathcal{F}\}$ *Calderón-Zygmund stopping data* for $f \in L_0^2(I_0)$, with constant C_{CZ} , if these properties hold.

- (1) I_0 is the maximal element of \mathcal{F} .
- (2) For all $I \in \mathcal{D}^\sigma$, $I \subset I_0$, we have $|\mathbb{E}_I^\sigma f| \leq C_{CZ} \alpha_f(\pi_{\mathcal{F}} I)$.

- (3) α_f is monotonic: If $F, F' \in \mathcal{F}$ and $F \subset F'$ then $\alpha_f(F) \geq \alpha_f(F')$.
- (4) The collection is σ -Carleson in the sense of (3.3), with constant C_{CZ} .
- (5) We have the inequality

$$(3.5) \quad \left\| \sum_{F \in \mathcal{F}} \alpha_f(F) \cdot F \right\|_{\sigma} \leq C_{CZ} \|f\|_{\sigma}.$$

We will consistently use the notation

$$P_F^{\sigma} f := \sum_{I \in \mathcal{D}^{\sigma} : \pi_{\mathcal{F}} I = F} \Delta_I^{\sigma} f.$$

We can fix the constant C_{CZ} to be some large fixed number for the remainder of the proof. We will very commonly derive sums of the form below, in which Q_F^w is some family of mutually orthogonal projections in $L^2(w)$.

$$\begin{aligned} \sum_{F \in \mathcal{F}} \{ \alpha_f(F) \sigma(F)^{1/2} + \|P_F^{\sigma} f\|_{\sigma} \} \|Q_F^w g\|_w \\ \leq \left[\sum_{F \in \mathcal{F}} \{ \alpha_f(F)^2 \sigma(F) + \|P_F^{\sigma} f\|_{\sigma}^2 \} \times \sum_{F \in \mathcal{F}} \|Q_F^w g\|_w^2 \right]^{1/2} \lesssim \|f\|_{\sigma} \|g\|_w. \end{aligned}$$

This follows from Cauchy-Schwarz and (3.5). This inequality we will refer to as the *quasi-orthogonality* argument. It is systemic to the proof.

The simplest way to select the stopping data is to take \mathcal{F} to be stopping intervals for the weighted averages of f . That is, if f is supported on interval I_0 , we construct the stopping data as follows. We set $I_0 \in \mathcal{F}$, defining $\alpha_f(I_0) = \mathbb{E}_{I_0}^{\sigma} |f|$. Inductively, for $F \in \mathcal{F}$, maximal subintervals $I \subset F$ such that $\mathbb{E}_I^{\sigma} |f| > 4\mathbb{E}_F^{\sigma} |f|$ is also in \mathcal{F} . We then take $\alpha_f(F) = \mathbb{E}_F^{\sigma} |f|$. This is Calderón-Zygmund stopping data for f with constant C_{CZ} bounded by an absolute constant. We will refer to this as the *standard Calderón-Zygmund stopping data*. There are however other choices that we will appeal to. The intervals \mathcal{F} need not be good intervals, and goodness of F will never be used in the proof.

Note that we do *not* assume that if $F, F' \in \mathcal{F}$, and $F' \subsetneq F$, then $\alpha_f(F') > C_0 \alpha_f(F)$. Indeed, for some applications, this property will not hold. This missing hypothesis is replaced by properties (3.3) and (3.5).

Let \mathcal{F}, α_f be Calderón-Zygmund stopping data for $f \in L_0^2(I_0, \sigma)$, and let \mathcal{G}, α_g be similar data for $g \in L_0^2(I_0, w)$. Define

$$(3.6) \quad B_{\mathcal{F}, \mathcal{G}}^{\text{near}}(f, g) := \sum_{\substack{(F, G) \in \mathcal{F} \times \mathcal{G} \\ \pi_{\mathcal{F}} G = F \text{ or } \pi_{\mathcal{G}} F = G}} B(P_F^{\sigma} f, P_G^w g).$$

This is a subtle definition. It is important to note that for fixed $F \in \mathcal{F}$, one can have many $G \in \mathcal{G}$ with \mathcal{F} -parent F . One should also note that if the stopping data for \mathcal{F} is in some sense ‘trivial’,

the bilinear form above is essentially indistinguishable from $\langle H_\sigma f, g \rangle_w$. On the other hand, the forms

$$\sum_{\substack{G \in \mathcal{G} \\ \pi_{\mathcal{F}} G = F}} \langle H_\sigma P_F^\sigma f, P_G^w g \rangle_w, \quad F \in \mathcal{F},$$

are much better, in that $P_F^\sigma f$ is more structured than an arbitrary $L^2(\sigma)$ function. This is one of the main results of this paper.

Theorem 3.7. *It holds that*

$$\left| B(f, g) - B_{\mathcal{F}, \mathcal{G}}^{\text{near}}(f, g) \right| \lesssim \mathcal{H} \|f\|_\sigma \|g\|_w.$$

We will refer to the application of this Theorem as the *parallel corona*. The proof of the Theorem depends upon the functional energy inequality in §7.

Proof. It is worth emphasizing that the presence of stopping data for *both* functions yields a seemingly essential reduction in complexity of the proof. Still, the proof requires some careful accounting of terms, with the fundamental fact needed for the proof being the functional energy inequality.

The inner product $\langle H_\sigma f, g \rangle_w$ is expanded as

$$\langle H_\sigma f, g \rangle_w = \sum_{F \in \mathcal{F}} \sum_{G \in \mathcal{G}} \langle H_\sigma P_F^\sigma f, P_G^w g \rangle_w.$$

The term $B_{\mathcal{F}, \mathcal{G}}^{\text{near}}(f, g)$ is a sum of pairs $(F, G) \in \mathcal{F} \times \mathcal{G}$ where either F is the minimal element of \mathcal{F} containing G , or the reverse statement holds. The complementary pairs (F, G) are those that are either disjoint, or $F \subset G$, but $G \in \mathcal{G}$ is not minimal with respect to inclusion, or the reverse statement holds. The difference between this and $\langle H_\sigma f, g \rangle_w$ is then the sum of the terms

$$\begin{aligned} B^{\text{disjoint}}(f, g) &:= \sum_{\substack{(F, G) \in \mathcal{F} \times \mathcal{G} \\ F \cap G = \emptyset}} \langle H_\sigma P_F^\sigma f, P_G^w g \rangle_w, \\ B^{\text{up}}(f, g) &:= \sum_{\substack{(F, G) \in \mathcal{F} \times \mathcal{G} \\ \pi_{\mathcal{F}} G \subsetneq F}} \langle H_\sigma P_F^\sigma f, P_G^w g \rangle_w, \end{aligned}$$

and a third form $B^{\text{down}}(f, g)$, which is dual the ‘up’ form, and so we do not explicitly address it. (We are dropping the subscripts \mathcal{F} and \mathcal{G} , as these will be fixed for the remainder of the argument.)

In the ‘disjoint’ form, we require $F \cap G = \emptyset$, so that it is the immediate corollary to (3.10) that there holds

$$\left| B^{\text{disjoint}}(f, g) \right| \lesssim \mathcal{H} \|f\|_\sigma \|g\|_w.$$

It remains to consider the form $B^{\text{up}}(f, g)$. In it, we require that $\pi_{\mathcal{F}} G \subsetneq F$.

It is the consequence of the elementary estimates (3.10) and (3.11) that for the up form it suffices to control the bilinear form

$$B(f, g) := \sum_{(I, J) : \pi_{\mathcal{F}}(\pi_{\mathcal{G}} J) \subseteq I} \mathbb{E}_{I_J}^{\sigma} \Delta_I^{\sigma} f \cdot \langle H_{\sigma} I_J, \Delta_J^w g \rangle_w.$$

Recall that I_J is the child of I that contains J .

For $F \in \mathcal{F}$, let

$$g_F = \sum_{J : \pi_{\mathcal{F}}(\pi_{\mathcal{G}} J) = F} \Delta_J^w g.$$

It is routine to check that these functions are $\mathcal{F}_{\mathbb{C}}$ -adapted. Therefore, the required estimate follows from Corollary 7.5. The proof is complete. \square

Remark 3.8. In the relevant literature, a corona refers to a decomposition of the bilinear form $B(f, g)$. Here, we are using the same term to indicate passage from a more general term to its essential part.

This section concludes with some standard facts in the subject. Define the collections of pairs of intervals

$$\begin{aligned} \mathcal{E} &:= \{(I, J) \in \mathcal{D}^{\sigma} \times \mathcal{D}^w : I \cap J = \emptyset \text{ or } 2^{-\rho}|J| \leq |I| \leq 2^{\rho}|J|\}, \\ \mathcal{E}_{\mathbb{C}} &:= \{(I, J) \in \mathcal{D}^{\sigma} \times \mathcal{D}^w : J \subsetneq I\}, \end{aligned}$$

and finally let \mathcal{E}_{\supset} be the collection of pairs of intervals dual to $\mathcal{E}_{\mathbb{C}}$. For the collection $\mathcal{E}_{\mathbb{C}}$, as $J \subsetneq I$, we have that J is contained in a child I_J of I .

Lemma 3.9. *We have the inequality below, for any integer $\rho \geq r$.*

$$(3.10) \quad \sum_{(I, J) \in \mathcal{E}} \left| \langle H_{\sigma} \Delta_I^{\sigma} f, \Delta_J^w g \rangle_w \right| \lesssim \mathcal{H} \|f\|_{\sigma} \|g\|_w.$$

The same inequality holds for the term below, and its dual, with $\mathcal{E}_{\mathbb{C}}$ replaced by \mathcal{E}_{\supset} , and the roles of w and σ reversed.

$$(3.11) \quad \sum_{(I, J) \in \mathcal{E}_{\mathbb{C}}} \left| \mathbb{E}_{I-I_J}^{\sigma} \Delta_I^{\sigma} f \cdot \langle H_{\sigma}(I - I_J), \Delta_J^w g \rangle_w \right|$$

In this expression, I_J is the child of I containing J .

This Lemma is implicit in [29]. The paper [9, Section 8] specifically points to this form of the Lemma; that section can be taken as a proof. We do not recall the proof.

4. A FUNCTION THEORY

In the first two sections, we build some function theory, focusing on the role of unbalanced Haar functions. We then apply that to the Hilbert transform in the third section. The function classes defined here are highly dependent upon the choice of dyadic grid, which is taken to be fixed. The main result of this section is Theorem 4.8, providing very sharp sufficient conditions for the two weight inequality, though in language specific to a dyadic grid, and testing the bilinear form over only functions that are good in that grid.

4.1. Bounded Fluctuation. The following class of functions extend the notion of being bounded, and are basic to our analysis.

Definition 4.1. A function f is of *bounded fluctuation on interval* I_0 , and we write $\|f\|_{b_1^\sigma(I_0)} \leq C$ if these two conditions hold.

- (1) $f \in L_0^2(I_0; \sigma)$.
- (2) On each dyadic interval $I \subset I_0$ on which f is not constant, it holds that $|\mathbb{E}_I^\sigma f| \leq C$.

We then take $\|f\|_{b_1^\sigma(I_0)}$ to be the best constant in this last condition, and also set

$$\|f\|_{B_1^\sigma(I_0)}^2 := \|f\|_{b_1^\sigma(I_0)}^2 \sigma(I_0) + \|f\|_\sigma^2.$$

The point of this definition is that a function f has two important quantities, the first is the number $\|f\|_{b_1^\sigma(I_0)}$, which is akin to the L^∞ norm of a bounded function. The second that the norm $\|f\|_{B_1^\sigma(I_0)}$ is motivated by the quasi-orthogonality considerations basic this paper. We will define more of these quasi-norms.

If we take standard Calderón-Zygmund stopping data for f , and f is balanced, then $\|P_F^\sigma f\|_\infty \lesssim \alpha_f(F)$, which is a final condition for us, in light of our hypotheses in the main theorem. But, if f is unbalanced, $P_F^\sigma f$ is only of bounded fluctuation with $\|f\|_{b_1^\sigma(F)} \lesssim \alpha_f(F)$. In both cases, we have

$$(4.2) \quad \sum_{F \in \mathcal{F}} \|P_F^\sigma f\|_{B_1^\sigma(F)}^2 \simeq \|f\|_\sigma^2.$$

For consistency of notation, let us set $\|f\|_{B_2^\sigma(I_0)}^2$ to be the infimum of the expressions $\sum_{F \in \mathcal{F}} \|f_F\|_{B_1^\sigma(F)}^2$, subject to the conditions $f = \sum_{F \in \mathcal{F}} f_F$, $P_F^\sigma f = f_F$, and \mathcal{F} is σ -Carleson. We have $\|f\|_{L_0^2(I_0, \sigma)} \simeq \|f\|_{B_2^\sigma(I_0)}$.

A set $\mathcal{Q} \subset \mathcal{D}$ is said to be *convex* if for all $I \subset I' \subset I''$, if $I, I'' \in \mathcal{Q}$, then so is I' . A projection $Q^\sigma = \sum_{I \in \mathcal{Q}} \Delta_I^\sigma$ associated with a convex subset of \mathcal{D} is a difference of two conditional expectation operators. In particular, if f is a bounded function, then so is Qf . This proposition is elementary, yet useful in our recursive application on the parallel corona.

Proposition 4.3. *Let \mathcal{Q} be a convex set, and let Q^σ denote the corresponding projection. Let f have Calderón-Zygmund stopping data \mathcal{F} and α_f . These properties hold.*

- (1) *If f is balanced, then $P_F^\sigma f$ is a bounded function, and in particular, $\|P_F^\sigma f\|_\infty \lesssim \alpha_f(F)$.*

- (2) If f is unbalanced, the function $P_F^\sigma f$ is of bounded fluctuation on F , with constant at most $2C_{CZ}\alpha_f(F)$. In particular, if K is a maximal interval with

$$(4.4) \quad |\mathbb{E}_K^\sigma P_F^\sigma f| > 2C_{CZ}\alpha_f(F),$$

then necessarily for each interval I in the Haar support of $P_F^\sigma f$, we have $K \subsetneq I$ or $I \cap K = \emptyset$. Moreover πK is in the Haar support of f . Here, C_0 is as in Definition 3.4.

- (3) Both of the assertions hold for $Q^\sigma P_F^\sigma f$, with constants multiplied by 2.

Proof. (1) The projection P_F^σ , defined at the end of Definition 3.4, is a convex partition. $P_F^\sigma f(x)$ is supported on F , and at each point $x \in F$, it is the difference

$$\mathbb{E}_K^\sigma f - \mathbb{E}_F^\sigma f,$$

where K is the smallest dyadic interval that contains x and has \mathcal{F} -parent F . Both averages are by assumption dominated in absolute value by $C_0\alpha_f(F)$, the key consequence of f being balanced. So the conclusion follows.

- (2) If f is unbalanced, the first part is just as in (1). If K is a maximal interval as in (4.4), then the \mathcal{F} parent of K can not be F . The conclusions then easily follow.
- (3) Convexity was the only property of P_F^σ used above, so the same conclusion will hold for $Q^\sigma P_F^\sigma f$.

□

4.2. Minimal Bounded Fluctuation. Functions of bounded fluctuation are still relatively unstructured, which necessitates this central refinement of the notion, namely *minimal bounded fluctuation*. We will refer to the notation established in the definition of balanced Haar functions, see Definition 3.1.

Definition 4.5. We say that f is a *minimal bounded fluctuation* function, and write $f \in \text{MBF}^\sigma(I_0)$, if f is in $B_1^\sigma(I_0)$, and if (1) f is unbalanced, (2) there is a collection \mathcal{K} of disjoint dyadic subintervals of I_0 , the *supporting intervals* of f , so that for each $K \in \mathcal{K}$, we have $K = (\pi K)_{\text{small}}$ and

$$\mathbb{E}_K^\sigma \Delta_{\pi K}^\sigma f < 0,$$

- (3) the Haar support of f equals $\pi\mathcal{K} := \{\pi K : K \in \mathcal{K}\}$, (so the Haar support of f is minimal) and (4) for each dyadic interval I not contained in some interval $K \in \mathcal{K}$, it holds that $0 \leq \mathbb{E}_I^\sigma f \leq \|f\|_{b_1^\sigma(I_0)}$. Note in particular, that this expectation equals a sum of positive quantities:

$$\mathbb{E}_I^\sigma f = \sum_{K \in \mathcal{K} : I \subset (\pi K)_{\text{large}}} \mathbb{E}_I^\sigma \Delta_{\pi K}^\sigma f.$$

We draw this conclusion: *Any Haar projection of a function of minimal bounded fluctuation is again of minimal bounded fluctuation.*

The need for the definition arises from the case where $\pi\mathcal{K}$ has substantial overlap. We take $\|f\|_{b_0^\sigma(I_0)}$ to be the best constant γ in a decomposition of $f = f_0 + f_1 \in L_0^2(I_0)$, where $\|f_0\|_\infty \leq \gamma$

and $f_1 \in \text{MBF}_0^\sigma(I_0)$ with $\|f_j\|_{b_1^\sigma(I_0)} \leq \gamma$, $j = 0, 1$. Define

$$\|f\|_{B_0^\sigma(I_0)}^2 := \|f\|_{b_0^\sigma(I_0)}^2 \sigma(I_0) + \|f\|_\sigma^2.$$

Keep in mind that the lower case letters in the norms indicate an L^∞ -like quantity, while capital letters indicate an L^2 -like quantity, that must ultimately be combined with a quasi-orthogonality argument.

We need a decomposition of a function of bounded fluctuation into a sum of bounded and minimal bounded fluctuation functions.

Lemma 4.6. *For any $0 < \epsilon < 1$, there is a constant C_ϵ so that this holds. Let $f \in B_1^\sigma(I_0)$ be unbalanced. There is an orthogonal Haar decomposition of f into functions $f = \phi_0 + \phi_1 + \phi_2$ so that these conditions hold.*

- (1) $\|\phi_2\|_\sigma \leq \epsilon \|f\|_{B_1^\sigma(I_0)}$. (Note the use of the $L^2(\sigma)$ norm here.)
- (2) The function ϕ_0 is bounded by $C_\epsilon \|f\|_{b_1^\sigma(I_0)}$.
- (3) The function ϕ_1 has Calderón–Zygmund stopping data \mathcal{F}_1 and $\alpha_{\phi_1}(\cdot)$ such that for each $F \in \mathcal{F}_1$, it holds that $P_F^\sigma \phi_1$ is of minimal bounded fluctuation and $\|P_F^\sigma \phi_1\|_{b_0^\sigma(I_0)} \lesssim \alpha_{\phi_1}(F)$.

Proof. We can assume that $\|f\|_{b_1^\sigma(I_0)} = 1$. Let \mathcal{K} be those intervals for which

$$\mathbb{E}_K^\sigma \Delta_{\pi K}^\sigma f < -4.$$

So, these are the dyadic intervals for which the martingale difference associated with the parent takes a large negative value on K . Observe that these intervals are pairwise disjoint, for if not, then for some interval $K \in \mathcal{K}$ we have that f is not constant on both K and πK . Then,

$$4 \leq |\mathbb{E}_K^\sigma \Delta_{\pi K}^\sigma f| = |\mathbb{E}_K^\sigma f - \mathbb{E}_{\pi K}^\sigma f| \leq 2.$$

And this is a contradiction.

Then, take $\phi_1 = \sum_{K \in \mathcal{K}} \Delta_{\pi K}^\sigma f$. Note that ϕ_1 satisfies most of the properties of the definition of minimal bounded fluctuation, but it certainly need not be of bounded fluctuation itself. Take \mathcal{L}_1 to be the maximal intervals in the collection $\pi\mathcal{K} := \{\pi K : K \in \mathcal{K}\}$. (Keep in mind that we need this Lemma in the case that \mathcal{L}_1 is much smaller than $\pi\mathcal{K}$.)

Then, for $L \in \mathcal{L}_1$, the function $\phi_1 \cdot L$ has standard Calderón–Zygmund stopping data $\mathcal{F}_{1,L}$, and $\alpha_{\phi_1,L}(\cdot)$. It is clear from the construction that each projection $P_F^\sigma(\phi_1 \cdot L)$ is a function of minimal bounded fluctuation, for $F \in \mathcal{F}_{1,L}$.

We can complete this to (non-standard) Calderón–Zygmund stopping data for ϕ_1 . Take $\mathcal{F}_1 := \{I_0\} \cup \bigcup_{L \in \mathcal{L}_1} \mathcal{F}_{1,L}$, and set $\alpha_{\phi_1}(I_0) := 0$, and otherwise takes the obvious value from the standard Calderón–Zygmund stopping data. It is evident from the construction that the third conclusion of the Lemma holds.

Let $\phi'_0 := f - \phi_1$. This is exactly the sort of function to which the conclusion of Lemma 4.7 applies. Namely, ϕ'_0 is a Haar projection of f onto those martingale differences Δ_I^σ with L^∞ -norm bounded by four. Hence, ϕ'_0 has exponential moments.

Choosing $C_\epsilon \simeq \log 1/\epsilon$, set $\phi'_0 = \phi_0 + \phi_2$, where

$$\phi_0 := \sum_{I: I \notin \{M\phi'_0 > C_\epsilon\}} \Delta_I^\sigma \phi'_0.$$

Then, it follows that ϕ_0 has L^∞ -norm at most $4 + C_\epsilon$. Meanwhile, $\|\phi_2\|_\sigma \lesssim \epsilon \sigma(I_0)^{1/2}$, which completes the proof. \square

There is a non-homogeneous variant of the the John-Nirenberg estimate.

Lemma 4.7. *Suppose that $f \in B_1^\sigma(I_0)$, and that \tilde{f} is any Haar projection of f for which $\|\Delta_I^\sigma \tilde{f}\|_\infty \leq 4\|f\|_{B_1^\sigma(I_0)}$, for all I . Then, we have the distributional estimate*

$$\sigma(|\tilde{f}| > C\lambda\|f\|_{B_1^\sigma(I_0)}) \leq \exp(-\lambda)\sigma(I_0), \quad \lambda > 1,$$

where C is an absolute constant.

Proof. Assume that $\|f\|_{B_1^\sigma(I_0)} = 1$. The fundamental properties are that (1) the martingale differences of \tilde{f} are bounded by 4, (2) the bounded fluctuation property of f , and (3) the universal weak-type $L^1(\sigma)$ for Haar projections.

We will show that there is an absolute constant $C > 4$ so that for $\lambda > 1$,

$$\sigma(|\tilde{f}| > \lambda + C) \leq e^{-1} \sigma(|\tilde{f}| > \lambda).$$

A standard induction on λ then completes the proof.

Let \mathcal{I}_λ be the maximal subintervals of I_0 such that $|\mathbb{E}_I^\sigma \tilde{f}| > \lambda$. It follows that the expectation cannot be more than $\lambda + 4$ is absolute value, by (1) above. Therefore,

$$\sigma(|\tilde{f}| > \lambda + C) \leq \sum_{I \in \mathcal{I}_\lambda} \sigma(\{x \in I : |\tilde{f} - \mathbb{E}_I^\sigma \tilde{f}| > C - 4\}).$$

If the set on the right is not empty, it follows that $\mathbb{E}_I^\sigma |f| \leq 1$ by (2) above. Now, we are only interested in the function $\tilde{f} - \mathbb{E}_I^\sigma \tilde{f}$ on the interval I , and it is a Haar projection of $f1_I$. This operation is weakly bounded on $L^1(I, \sigma)$, which is (3) above. Hence,

$$\sigma(|\tilde{f} - \mathbb{E}_I^\sigma \tilde{f}| > C - 4) \lesssim (C - 4)^{-1} \|f1_I\|_{L^1(w)} \lesssim (C - 4)^{-1} \sigma(I).$$

This clearly completes the proof. \square

4.3. Sufficient Conditions for the Two Weight Inequality. This is very nearly a characterization of the two weight inequality, and an essential intermediate step in our goal.

Theorem 4.8. *Suppose that the pair of weights (σ, w) do not share common point mass, and fix a dyadic grid \mathcal{D} . Consider the two weight inequality*

$$(4.9) \quad |\langle H_\sigma f, g \rangle_w| \leq \mathcal{N}_{\text{good}} \|f\|_\sigma \|g\|_w,$$

where it is required that $f = P_{\text{good}}^\sigma f \in L^2(\sigma)$ and $g = P_{\text{good}}^w g \in L^2(w)$. It holds that $\mathcal{N}_{\text{good}} \lesssim \mathcal{H} + \mathcal{B}_0$, where the latter constant is the best constant in the inequalities below, for $I \in \mathcal{D}$

$$\int_I |H_\sigma f|^2 dw \leq \mathcal{B}_0^2 \|f\|_{B_0^\sigma(I)}^2, \quad \int_I |H_w g|^2 d\sigma \leq \mathcal{B}_0^2 \|g\|_{B_0^w(I)}^2.$$

Note that goodness depends upon the grid, which is taken fixed in (4.9). Certainly the testing conditions above are necessary for the boundedness of the bilinear form. We have not sought to show that the A_2 condition is necessary from the condition (4.9). But, it is a consequence of the random grid construction that $\mathcal{N} \lesssim \mathbb{E}_{\mathcal{D}} \mathcal{N}_{\text{good}}$, where the expectation over grids is taken appropriately, for details on this see [8, 29].

The Theorem is a consequence of the parallel corona. Using our definition of the function classes above, define several constants $\mathcal{N}_{i,j}$, $0 \leq i \leq j \leq 2$ to be the best constant in the inequalities

$$\begin{aligned} |\langle H_\sigma f, g \rangle_w| &\leq \mathcal{N}_{i,j} \|f\|_{B_i^\sigma(I)} \|g\|_w, & g \in B_j^w(I), \quad f \text{ is good}, \\ |\langle H_\sigma f, g \rangle_w| &\leq \mathcal{N}_{i,j} \|f\|_\sigma \|g\|_{B_i^w(I)}, & f \in B_j^\sigma(I), \quad g \text{ is good}. \end{aligned}$$

Note that the right hand side places an L^2 norm on the function in the larger function class. An important consequence of the parallel corona is

Theorem 4.10. *Let $1 \leq i \leq j \leq 2$. There holds*

$$\mathcal{N}_{i,j} \lesssim \begin{cases} \mathcal{H} + \mathcal{N}_{i-1,j} + \mathcal{N}_{i,j-1} & i < j \\ \mathcal{H} + \mathcal{N}_{i-1,j} & i = j \end{cases}.$$

This gives the following inequalities, by recursive application.

$$\begin{aligned} \mathcal{N}_{\text{good}} &\simeq \mathcal{N}_{2,2} \lesssim \mathcal{H} + \mathcal{N}_{1,2} \\ &\lesssim \mathcal{H} + \mathcal{N}_{0,2} + \mathcal{N}_{1,1} \\ &\lesssim \mathcal{H} + \mathcal{N}_{0,2} + \mathcal{N}_{0,1} \simeq \mathcal{H} + \mathcal{N}_{0,2} \end{aligned}$$

The conditions $\mathcal{N}_{0,2}$ are the testing hypothesis on functions in the classes $B_0^\sigma(I)$ and $B_0^w(I)$, and trivially, $\mathcal{N}_{0,1} \leq \mathcal{N}_{0,2}$. In §6 we will show that testing over MBF functions is controlled by the A_2 and the L^∞ testing constant, completing our proof of the main Theorem.

Proof. The basic tools here are (4.2), the parallel corona, and the more complicated Lemma 4.6. First, an easy case, namely $\mathcal{N}_{2,2} \lesssim \mathcal{H} + \mathcal{N}_{1,2}$. Indeed, taking $f \in L^2(\sigma)$ and $g \in L^2(w)$, we can take Calderón–Zygmund stopping data \mathcal{F} and $\alpha_f(\cdot)$ for f and \mathcal{G} and $\alpha_g(\cdot)$ for g . Then, applying the parallel corona leads immediately to

$$|\langle H_\sigma f, g \rangle_w| \lesssim \mathcal{H} \|f\|_\sigma \|g\|_w + |\mathcal{B}_{\mathcal{F}, \mathcal{G}}^{\text{near}}(f, g)|.$$

Recall that the ‘near’ form is defined in (3.6). Since $P_F^\sigma f \in B_1^\sigma(F)$, for each $F \in \mathcal{F}$, we have

$$\sum_{F \in \mathcal{F}} \left| \sum_{G \in \mathcal{G} : \pi_{\mathcal{F}} G = F} \langle H_\sigma P_F^\sigma f, P_G^w g \rangle_w \right| \leq \mathcal{N}_{1,2} \sum_{F \in \mathcal{F}} \|P_F^\sigma f\|_{B_1^\sigma(F)} \left\| \sum_{G \in \mathcal{G} : \pi_{\mathcal{F}} G = F} P_G^w g \right\|_w$$

$$\lesssim \mathcal{N}_{1,2} \|f\|_{\sigma} \|g\|_w,$$

where the last line follows from quasi-orthogonality (4.2). This completes the proof.

The arguments involving $\mathcal{N}_{i,j}$ with i or j equal to one are more complicated, due to the statement of Lemma 4.6, and in particular, the fact that the absorption term associated with ϕ_2 involves $L^2(\sigma)$. For the sake of specificity, let us argue that $\mathcal{N}_{1,2} \lesssim \mathcal{H} + \mathcal{N}_{1,1} + \mathcal{N}_{0,2}$. All other cases are similar.

Select interval I_0 , and f, g so that

$$(4.11) \quad \frac{7}{8} \mathcal{N}_{1,2} \|f\|_{B_1^{\sigma}(I_0)} \|g\|_w \leq \langle H_{\sigma} f, g \rangle_w.$$

Applying Lemma 4.6 to f , with a choice of ϵ to be specified, we have $f = f_0 + f_1 + f_2$, where f_0 is a function bounded by C_{ϵ} , and f_1 have non-trivial Calderón-Zygmund stopping data \mathcal{F}_1 and $\alpha_{f_1}(\cdot)$, so that

$$\|f\|_{B_1^{\sigma}(I_0)}^2 \gtrsim \sum_{F \in \mathcal{F}_1} \|P_F^{\sigma} f_1\|_{B_0^{\sigma}(F)}^2.$$

For f_2 , there holds $\|f_2\|_{\sigma} \leq \epsilon \|f\|_{B_1^{\sigma}(I_0)}$. On the other hand, we take standard Calderón-Zygmund stopping data \mathcal{G} and $\alpha_g(\cdot)$ for g .

Note that for the terms that involve f_2 and g we have

$$|\langle H_{\sigma} f_2, g \rangle_w| \leq \epsilon \mathcal{N}_{2,2} \|f_2\|_{\sigma} \|g\|_w \lesssim \epsilon \{\mathcal{H} + \mathcal{N}_{1,2}\} \|f\|_{\sigma} \|g\|_w$$

by the first step in the proof. Thus, this term can be absorbed into the left hand side of (4.11), for ϵ sufficiently small.

The main term is then $\langle H_{\sigma}(f_0 + f_1), g \rangle_w$. For the function f_0 , we just use the L^{∞} -bound, and the constant $\mathcal{N}_{0,2}$.

$$|\langle H_{\sigma} f_0, g \rangle_w| \leq C_{\epsilon} \mathcal{N}_{0,2} \sigma(I_0)^{1/2} \|g\|_w.$$

That leaves the function f_1 . We have stopping data for both f_1 and g , so that repeating the argument involving the parallel corona and the quasi-orthogonality, we see that

$$|\langle H_{\sigma} f_1, g \rangle_w| \lesssim \{\mathcal{H} + \mathcal{N}_{0,2} + \mathcal{N}_{1,1}\} \|f\|_{B_1^{\sigma}(I_0)} \|g\|_w.$$

Collecting estimates, we have shown that

$$\frac{7}{8} \mathcal{N}_{2,1} \lesssim \mathcal{H} + \epsilon \mathcal{N}_{2,1} + C_{\epsilon} \mathcal{N}_{0,2} + \mathcal{N}_{1,1}.$$

As this holds for all $0 < \epsilon < 1$, this completes the proof of $\mathcal{N}_{2,1} \lesssim \mathcal{H} + \mathcal{N}_{1,1}$.

The estimate for $\mathcal{N}_{1,1}$ is similar to those presented.

□

5. ENERGY, MONOTONICITY, AND POISSON

Our Theorem is particular to the Hilbert transform, and so depends upon special properties of it. They largely extend from the fact that the derivative of $-1/y$ is positive. The following Monotonicity Property for the Hilbert transform was observed in [9, Lemma 5.8], and is basic to the analysis of the functional energy inequality. We will frequently use the notation $J \Subset I$ to mean $J \subset I$ and $2^*|J| \leq |I|$, so that the property of J being good becomes available to us.

Lemma 5.1 (Monotonicity Property). *Suppose that ν is a signed measure, and μ is a positive measure with $\mu \geq |\nu|$, both supported outside an interval $I \in \mathcal{D}^\sigma$. Then, for good $J \Subset I$, and function $g \in L_0^2(J, w)$, it holds that*

$$(5.2) \quad |\langle H\nu, g \rangle_w| \leq \langle H\mu, \bar{g} \rangle_w \approx P(J, \mu) \left\langle \frac{x}{|J|}, \bar{g} \right\rangle_w.$$

Here, $\bar{g} = \sum_{J'} |\hat{g}(J')| h_{J'}^w$, is a Haar multiplier applied to g .

Proof. By linearity, it suffices to consider the case of $g(x) = h_J^w(x)$. Namely, we should prove

$$|\langle H\nu, h_J^w \rangle_w| \leq \langle H\mu, h_J^w \rangle_w \approx P(J, \mu) \left\langle \frac{x}{|J|}, h_J^w \right\rangle_w.$$

The function $H\mu$ is monotonically increasing on I , and we have defined the Haar functions so that the inner product $\langle H\mu, h_J^w \rangle_w$ is positive, as is $\langle x, h_J^w \rangle_w$. So the right hand side above is non-negative.

The Haar function h_J^w is also constant on both halves of J , and has w -integral zero. Thus, there is a monotonic increasing map $\phi : J_+ \rightarrow J_-$ so that

$$\langle H\nu, h_J^w \rangle_w = \int_{J_+} (-H\nu(x) + H\nu(\phi(x))) h_J^w(x) dw$$

The Haar function is positive on J_+ . Under the assumption that $|\nu| \leq \mu$, we make the difference on the right both bigger in absolute value, and positive, by replacing $H\nu$ with $H\mu$. This proves the first inequality.

To compare to the Poisson integral, we examine the inner product $\langle H\mu, h_J^w \rangle_w$. Let us write, for $x \in J$ and $y \in \mathbb{R} - I$, and x_J the center of J ,

$$\begin{aligned} \frac{1}{y-x} &= \frac{1}{(y-x_J) - (x-x_J)} \\ &= \frac{1}{y-x_J} \cdot \frac{1}{1 - \frac{x_J-x}{y-x_J}} = \frac{1}{y-x_J} \sum_{k=0}^{\infty} \left[\frac{x_J-x}{y-x_J} \right]^k. \end{aligned}$$

Therefore, it follows that

$$\begin{aligned} \langle H\mu, h_J^w \rangle_w &= \langle H\mu - H\mu(x_J), h_J^w \rangle_w \\ &= \int_{\mathbb{R}-I} \int_J \left\{ \frac{1}{x-y} - \frac{1}{y-x_J} \right\} h_J^w(x) dw(x) d\mu(y) \end{aligned}$$

$$= \sum_{k=1}^{\infty} \int_{\mathbb{R}-I} \int_J \frac{(x_J - x)^k}{(y - x_J)^{k+1}} h_J^w(x) dw(x) d\mu(y)$$

Recall that the condition that J be good implies that $\text{dist}(\partial I, J) \geq 2^{(1-\epsilon)r}|J|$. The term $k = 1$ is

$$cP(|\mu|, J) \left\langle \frac{x}{|J|}, h_J^w \right\rangle_w.$$

where $|c - 1| \lesssim 2^{(\epsilon-1)r}$. All of the higher order terms are geometrically less than this. For $k \geq 2$, note that we will always have

$$\left| \int_{\mathbb{R}-I} \frac{|J|^k}{(y - x_J)^{k+1}} d\mu \right| \lesssim 2^{(\epsilon-1)r(k-1)} \int_{\mathbb{R}-I} \frac{|J|}{(y - x_J)^2} d\mu.$$

And critically, by examination,

$$\left| \frac{(x - x_J)^k}{|J|^k} h_J^w(x) \right| \leq 2^{-k+1} \frac{x - x_J}{|J|} h_J^w(x) \quad x \in J, \quad k \geq 2.$$

□

The concept of *energy* is fundamental to the subject. For interval I , define

$$E(w, I)^2 := \mathbb{E}_I^{w(dx)} \mathbb{E}_I^{w(dx')} \frac{(x - x')^2}{|I|^2}.$$

Now, consider the *energy constant*, the smallest constant \mathcal{E} such that this condition holds, as presented or in its dual formulation. For all intervals I_0 , all partitions \mathcal{P} of I , it holds that

$$(5.3) \quad \sum_{I \in \mathcal{P}} P(\sigma I_0, I)^2 E(w, I)^2 w(I) \leq \mathcal{E}^2 \sigma(I_0).$$

It is was shown in [8, Proposition 2.11], that $\mathcal{E} \lesssim \mathcal{A}_2^{1/2} + \mathcal{T} = \mathcal{H}$. We will always estimate \mathcal{E} by \mathcal{H} .

One should keep in mind that the concept of energy is related to the tails of the Hilbert transform. The energy inequality, and its multi-scale extension to the functional energy inequality, show that the control of the tails is very subtle in this problem.

We also need the following elementary Poisson estimate from [29]; used occasionally in this argument, it is crucial to the proofs of Lemma 3.9 and (3.11).

Lemma 5.4. *Suppose that $J \Subset I \subset \widehat{I}$, and that J is good. Then*

$$(5.5) \quad |J|^{2\epsilon-1} P(\sigma(\widehat{I} - I), J) \lesssim |I|^{2\epsilon-1} P(\sigma(\widehat{I} - I), I).$$

Proof. We have $\text{dist}(J, \widehat{I} - I) \geq |J|^\epsilon |I|^{1-\epsilon}$, so that for any $x \in \widehat{I} - I$, we have

$$\frac{1}{(|J| + \text{dist}(x, J))^2} \lesssim \left[\frac{|I|}{|J|} \right]^{2\epsilon} \frac{1}{(|I| + \text{dist}(x, I))^2}$$

Hence, we have

$$\begin{aligned} |J|^{2\epsilon-1} P(\sigma \cdot (\hat{I} - I), J) &= |J|^{2\epsilon-1} \int_{\hat{I}-I} \frac{|J|}{(|J| + \text{dist}(x, J))^2} d\sigma \\ &\lesssim |I|^{2\epsilon} \int_{\hat{I}-I} \frac{1}{(|J| + \text{dist}(x, J))^2} d\sigma. \end{aligned}$$

And this proves the inequality. \square

6. TESTING MINIMAL BOUNDED FLUCTUATION FUNCTIONS

In Theorem 4.8, we have reduced the two weight inequality to testing of functions in the class $B_0^\sigma(I_0)$, and the dual collection, in any fixed dyadic grid \mathcal{D} . These functions are the linear combinations of bounded functions and those of minimal bounded fluctuation. In this section, we show that testing of functions of minimal bounded fluctuation follows from the A_2 condition and testing of bounded functions. As a corollary, testing bounded functions is sufficient for the two weight inequality, which is the main result of this paper.

Theorem 6.1. *The inequality below and its dual hold. For all intervals I and $f \in B_0^\sigma(I_0) \cap \text{MBF}^\sigma(I_0)$,*

$$\int_I |H_\sigma f|^2 dw \lesssim \{\mathcal{A}_2^{1/2} + \mathcal{T}_\infty\} \|f\|_{B_0^\sigma(I_0)}^2.$$

The very specific structure of functions in the class MBF^σ is essential to this argument, and even still, the argument is intricate. An outline of the arguments is as follows.

- (1) There are some initial general steps, the first being the classical above and below splitting, common to most T1-type arguments.
- (2) The second is a corona argument, another consequence of functional energy. This allows essential simplifications. One that we have not seen to date is the use of the *energy corona*, which smooths out certain irregularities of the pair of weights. This argument originates in [17], but is herein implemented under conditions which are necessary from the A_2 and testing hypotheses.
- (3) The third is the passage to the *stopping terms*, the name coming from [17, 29], see (6.11). This argument has been used before, especially the proof of Corollary 7.5. There however, one was ‘stopping’ at intervals in Calderón-Zygmund stopping data. Now, the ‘stopping’ intervals are not sparse, and our difficulties multiply.
- (4) After this, the argument comes in two parts, depending upon the position of the MBF function.

The general steps begin with the above and below splitting of the bilinear form $\langle H_\sigma f, g \rangle_w$.

Corollary 6.2. *Let $f \in L_0^2(I_0, \sigma)$ and $g \in L_0^2(I_0, w)$. It holds that*

$$\left| \langle H_\sigma f, g \rangle_w - \{B^{\text{above}}(f, g) + B^{\text{below}}(f, g)\} \right| \lesssim \mathcal{H} \|f\|_\sigma \|g\|_w$$

where

$$B^{\text{above}}(f, g) := \sum_{I \subset I_0} \sum_{J : J \in I} \mathbb{E}_{I_J} \Delta_I^\sigma f \cdot \langle H_\sigma I_J, \Delta_J^w g \rangle_w$$

and $B^{\text{below}}(f, g)$ is defined in the dual fashion.

This is a corollary to the uncomplicated estimates (3.10) and (3.11). It is perhaps a useful remark that all prior arguments on this question have made the step above at an early stage of the argument. Our main theorem cannot be proved in this fashion, hence our use of the parallel corona.

To establish Theorem 6.1, it suffices to establish this proposition. Note that the L^∞ testing constant appears in the first estimate.

Proposition 6.3. *The two inequalities below, and their duals, hold. For all intervals I_0 ,*

$$(6.4) \quad |B^{\text{above}}(f, g)| \lesssim \{\mathcal{A}_2^{1/2} + \mathcal{T}_\infty\} \|f\|_\sigma \|g\|_{B_0^w(I_0)}, \quad g \in \text{MBF}^w(I_0),$$

$$(6.5) \quad |B^{\text{above}}(f, g)| \lesssim \mathcal{H} \|f\|_{B_0^\sigma(I_0)} \|g\|_w, \quad f \in \text{MBF}^\sigma(I_0).$$

In the form $B^{\text{above}}(f, g)$, we will say that f is in the *up position*, and g is in the *down position*. Our proof splits according to the position of the MBF function.

We need a notion of a corona. Let \mathcal{F} , $\alpha_f(\cdot)$ be Calderón-Zygmund stopping data for f . Set P_F^σ as in Definition 3.4, and set $Q_F^w g := \sum_{J : \pi_{\mathcal{F}} J = F} \Delta_J^w g$ be the same projection into $L^2(w)$. Define

$$B_{\mathcal{F}}^{\text{above}}(f, g) := \sum_{F \in \mathcal{F}} B^{\text{above}}(P_F^\sigma f, Q_F^w g).$$

The corona estimate is

Theorem 6.6. *There holds*

$$|B^{\text{above}}(f, g) - B_{\mathcal{F}}^{\text{above}}(f, g)| \lesssim \mathcal{H} \|f\|_\sigma \|g\|_w.$$

It is an important complication that we do not know of any variant of this estimate for stopping data on the down function. As a consequence, in order to use the corona estimate, it is required that we have the $L^2(w)$ norm on g .

Proof. We apply functional energy. Expand

$$B^{\text{above}}(f, g) = \sum_{F' \in \mathcal{F}} \sum_{F \in \mathcal{F}} B^{\text{above}}(P_{F'}^\sigma f, Q_F^w g)$$

In the sum above, we can also add the restriction that $F' \supset F$. The case of $F' = F$ is the definition of $B_{\mathcal{F}}^{\text{above}}(f, g)$, so that it suffices to estimate

$$\sum_{\substack{F, F' \in \mathcal{F} \\ F' \supsetneq F}} B^{\text{above}}(P_{F'}^\sigma f, Q_F^w g).$$

Observe that the functions

$$g_F := \sum_{\substack{J: J \in F \\ \pi_{\mathcal{F}} J = F}} \Delta_J^w g$$

are \mathcal{F}_C adapted in the sense of Definition 7.1. Therefore, the Theorem is the corollary to Corollary 7.5. The proof is complete. \square

There is a general construction, the *energy corona*, which structures the pair of weights in a significant way.

Definition 6.7. Given interval $I_1 \subset I_0$, define $\mathcal{F}_{\text{energy}}(I_1)$ to be the maximal subintervals $I \subsetneq I_1$ such that there is a partition $\mathcal{J}(I)$ of I into intervals good intervals $J \in I$ with

$$(6.8) \quad \sum_{J \in \mathcal{J}(I)} P(\sigma I_1, J)^2 E(w, J)^2 w(J) > 10 \mathcal{E}^2 \sigma(I).$$

Here, \mathcal{E} is the constant in (5.3), and it holds that $\mathcal{E} \lesssim \mathcal{H}$.

We then set $\mathcal{F} := \bigcup_{n=0}^{\infty} \mathcal{F}_n$, where $\mathcal{F}_0 := \{I_0\}$, and inductively set $\mathcal{F}_{n+1} := \bigcup_{I \in \mathcal{F}_n} \mathcal{F}_{\text{energy}}(I)$. These are the *energy stopping intervals*.

The fact that the energy inequality (5.3) holds shows that \mathcal{F} is σ -Carleson. Let $\mathcal{M}, \alpha_f(\cdot)$ be standard Calderón-Zygmund stopping data for f , and then extend α_f to \mathcal{F} by setting $\alpha_f(F) = \alpha_f(\pi_{\mathcal{M}} F)$. Then, $\mathcal{F} \cup \mathcal{M}$, with the extended function α_f is Calderón-Zygmund stopping data for f . We will refer to this as the energy stopping data for f . (It is this step that requires our general definition of stopping data.)

This definition will be useful to summarize a frequent hypothesis in lemmas below.

Definition 6.9. Say that $f \in B_0^\sigma(I_0)$ is *energy-regular* if the condition that I is contained in some $I' \in \mathcal{F}_{\text{energy}}(I_0)$ implies that $\hat{f}(I) = 0$.

By the construction of the energy corona, and a straight forward application of Theorem 6.6, it suffices to assume that f is energy-regular below.

Remark 6.10. The considerations above have their origins in [17, 29], although they are used here with the necessary energy inequality for the first time.

The *stopping term* will be the main focus of attention. Let $f \in B_0^\sigma(I_0)$, with $\|f\|_{B_0^\sigma(I_0)} = 1$. Further assume, as will be the case below, that if K is an interval on which f takes constant value greater than one in absolute value, then g is constant on that interval. Write the argument of the Hilbert transform as $I_J = I_0 - (I_0 - I_J)$. This yields a decomposition of the ‘above’ form into two, the first is

$$\sum_{I: I \subset I_0} \sum_{J: J \in I} \mathbb{E}_{I_J} \Delta_I^\sigma f \cdot \langle H_\sigma I_0, \Delta_J^w g \rangle_w \lesssim \|f\|_{B_0^\sigma(I_0)} \left| \langle H_\sigma I_0, g \rangle_w \right|$$

$$\lesssim \mathcal{H} \|f\|_{b_0^\sigma(I_0)} \sigma(I_0)^{1/2} \|g\|_w,$$

by the fact that f is of bounded fluctuation, and the interval testing condition. The second term is the *stopping term*:

$$(6.11) \quad B^{\text{stop}}(f, g) := \sum_{I: I \subset I_0} \sum_{J: J \in I} \mathbb{E}_{I_J} \Delta_I^\sigma f \cdot \langle H_\sigma(I_0 - I_J), \Delta_J^w g \rangle_w$$

We will make the standing assumption that $g = \sum_J |\hat{g}(J)| h_J^w$, which maximizes the relevant inner products in the stopping term. (Compare this to the proof of Corollary 7.5.)

6.1. Minimal Bounded Fluctuation in the Up Position. We concern ourselves with the proof of the estimate (6.5), under the assumption that f is energy-regular.

With an arbitrary L^2 function in the down position, we have only a small number of tools to control the stopping term. And so the proof turns on the very strong ‘positivity’ property of MBF functions noted already in Definition 4.5.

Take $f \in \text{MBF}^\sigma(I_0)$, with $\|f\|_{b_0^\sigma(I_0)} = 1$, supporting intervals \mathcal{K} , which are the pairwise disjoint intervals on which $\Delta_{\pi K}^\sigma f$ takes a negative value. Moreover the Haar support of f equals $\pi K := \{\pi K : K \in \mathcal{K}\}$.

First, we can assume that g is constant on each interval $K \in \mathcal{K}$. To see this, set $g_K := \sum_{J: J \in K} \Delta_J^w g$. Using Lemma 5.1,

$$|B^{\text{stop}}(f, g_K)| \lesssim P(\sigma, K) \left\langle \frac{\chi}{|K|}, g_K \right\rangle_w.$$

The intervals \mathcal{K} are pairwise disjoint, so that the sum over $K \in \mathcal{K}$ of this last expression is controlled by Cauchy-Schwarz and the energy inequality (5.3). We can therefore assume that $g_K \equiv 0$ for all K , proving our claim.

Second, the essential point is then this positivity property: For each interval $I \in \pi K$, and J in the Haar support of g , we have $\mathbb{E}_{I_J}^\sigma \Delta_I^\sigma f > 0$, and moreover

$$0 < \sum_{I: I \supset J} \mathbb{E}_{I_J}^\sigma \Delta_I^\sigma f \leq 1.$$

Let $\mathcal{P}_0 := \{(I, J) : J \in I, \hat{f}(I) \neq 0, \hat{g}(J) \neq 0\}$. Given $\mathcal{P} \subset \mathcal{P}_0$, make these definitions. The first is a restriction of the stopping term, define

$$B_{\mathcal{P}}^{\text{stop}}(f, g) := \sum_{(I, J) \in \mathcal{P}} \mathbb{E}_{I_J}^\sigma \Delta_I^\sigma f \cdot \langle H_\sigma(I_0 - I_J), \Delta_J^w g \rangle_w.$$

Also, define a notion of size as follows.

$$\text{size}(\mathcal{P}) := \sup_{I_1: I_1 \subset I_0} \sum_{\substack{I: \exists (I, J) \in \mathcal{P} \\ J \subset I_1 \subset I}} \mathbb{E}_{I_J}^\sigma \Delta_I^\sigma f,$$

It is clear that $\text{size}(\mathcal{P}_0) \leq 1$. The main Lemma is as follows.

Lemma 6.12. *A collection $\mathcal{P} \subset \mathcal{P}_0$ can be decomposed into $\mathcal{P}_{\text{big}} \cup \mathcal{P}_{\text{small}}$ so that*

$$\begin{aligned} B_{\mathcal{P}}^{\text{stop}}(f, g) &\lesssim \text{size}(\mathcal{P}) \mathcal{H}\sigma(I_0)^{1/2} \|g\|_w, \\ \text{size}(\mathcal{P}_{\text{small}}) &\leq \frac{3}{4} \text{size}(\mathcal{P}). \end{aligned}$$

It is clear that this Lemma concludes the proof. For starting at the collection \mathcal{P}_0 , we can apply it recursively to gain a decomposition $\mathcal{P}_0 = \bigcup_{m \geq 1} \mathcal{P}_m$ with

$$B_{\mathcal{P}_m}^{\text{stop}}(f, g) \lesssim \left(\frac{3}{4}\right)^m \mathcal{H}\sigma(I_0)^{1/2} \|g\|_w.$$

This is clearly summable in $m \geq 1$ to the bound we want.

Proof. Let $\tau = \text{size}(\mathcal{P})$. The main recursion consists of constructing a collection \mathcal{I} of disjoint intervals. Initialize $\mathcal{Q} \leftarrow \mathcal{P}$, $\mathcal{P}_{\text{big}} \leftarrow \emptyset$, and $\mathcal{I} \leftarrow \emptyset$. While it holds that

$$\text{size}(\mathcal{Q}) \geq \frac{3}{4} \tau,$$

consider intervals I_1 which satisfy

$$\sum_{\substack{I: \exists (I, J) \in \mathcal{P} \\ J \subset I_1 \subset I}} \mathbb{E}_{I_1}^\sigma \Delta_I^\sigma f \geq \frac{3}{4} \tau.$$

Take I_1 maximal with respect to inclusion, and from these, select one with $\sigma(I_1)$ maximal. Then, define $\mathcal{P}_{I_1} := \{(I, J) \in \mathcal{Q} : I \supset I_1 \supset J\}$, and update

$$\mathcal{I} \leftarrow \mathcal{I} \cup \{I_1\}, \quad \mathcal{P}_{\text{big}} \leftarrow \mathcal{P}_{\text{big}} \cup \mathcal{P}_{I_1}, \quad \mathcal{Q} \leftarrow \mathcal{Q} - \mathcal{P}_{I_1}.$$

The intervals in \mathcal{I} are pairwise disjoint. For contradiction, assume that $I_2 \subsetneq I_1$ are both in \mathcal{I} . Then I_1 was selected for membership in \mathcal{I} first, and all pairs (I, J) with $J \subset I_1 \subset I$ were added to \mathcal{P}_{I_1} . Hence, for all $(I, J) \in \mathcal{P}_{I_2}$, it holds that $I \subsetneq I_1$. By definition,

$$\sum_{v=1}^2 \sum_{\substack{I: \exists (I, J) \in \mathcal{P}_{I_v} \\ J \subset I_v \subset I}} \mathbb{E}_{I_v}^\sigma \Delta_I^\sigma f \leq \tau,$$

yet for both $v = 1, 2$ the second sum exceeds $\frac{3}{4} \tau$, which is a contradiction.

We use the monotonicity property (5.1), and the energy-regularity of f , whence (6.8) fails that for each $I \in \mathcal{I}$ with $\hat{f}(I) \neq 0$. These facts give us

$$B_{\mathcal{P}_I}^{\text{stop}}(f, g) \lesssim \tau \mathcal{H}\sigma(I)^{1/2} \|g_I\|_w$$

where $g_I := \sum_{J: J \subset I} \Delta_J^w g$. Disjointness of the intervals I and Cauchy-Schwarz conclude the proof of the Lemma. \square

Remark 6.13. In order to verify the Nazarov-Treil-Volberg conjecture, one should show that, if $f \in B_0^\sigma(I_0)$, and f is energy-regular, then

$$|B^{\text{stop}}(f, g)| \lesssim \mathcal{H}\|f\|_{B_0^\sigma(I_0)} \|g\|_w.$$

But, in the argument above, we have relied upon the fact that the partial sums of the martingale differences of an MBF function have finite variation. A bounded function need not even have

martingale differences with finite *quadratic* variation, so there is no clear generalization of the proof above.

6.2. Minimal Bounded Fluctuation in the Down Position. The proof of the estimate (6.4) is reduced to L^∞ testing and the already established estimate (6.5). The corona is essential. The function f in the up position is assumed to be in $L^2_\sigma(I_0)$. By first applying Theorem 6.6 with standard Calderón-Zygmund stopping data for f , it follows that it is sufficient to prove

$$|B^{\text{above}}(f, g)| \lesssim \mathcal{H} \|f\|_{B^\sigma_1(I_0)} \|g\|_w, \quad g \in \text{MBF}^w(I_0).$$

The $L^2(\sigma)$ norm has been replaced by $B^\sigma_1(I_0)$, and the $B^w_0(I_0)$ norm has been replaced by the smaller $L^2(w)$ norm. By using the decomposition of Lemma 4.6, it suffices to consider the inequality above for f in the smaller class $B^\sigma_0(I_0)$, with the corresponding norm on the right hand side. Now, the class $B^\sigma_0(I_0)$ is the finite linear combination of bounded functions and functions in the class $\text{MBF}^\sigma(I_0)$. The case of a minimal bounded function in the up position is the estimate of (6.5). Therefore, the remaining estimate to prove is as below, in which we invoke the L^∞ testing constant.

Lemma 6.14. *For an interval I_0 , functions $f \in L^\infty(\sigma, I_0)$ and $g \in \text{MBF}^w(I_0)$, there holds*

$$|B^{\text{above}}(f, g)| \lesssim \{\mathcal{A}_\infty^{1/2} + \mathcal{T}_\infty\} \|f\|_\infty \sigma(I_0)^{1/2} \|g\|_w.$$

Proof. Note that we have by assumption on L^∞ testing,

$$|\langle H_\sigma f, g \rangle_w| \leq \mathcal{T}_\infty \|f\|_\infty \sigma(I_0)^{1/2} \|g\|_w.$$

Starting from this inequality, we also have, by Corollary 6.2,

$$|\langle H_\sigma f, g \rangle_w - \{B^{\text{above}}(f, g) + B^{\text{below}}(f, g)\}| \lesssim \mathcal{H} \|f\|_\sigma \|g\|_w.$$

The right hand side is smaller than our first estimate. Moreover, the assumption that $g \in \text{MBF}^w(I_0)$, and the estimate (6.5) for when the minimal bounded fluctuation term in the up position give

$$(6.15) \quad |B^{\text{below}}(f, g)| \lesssim \mathcal{H} \|f\|_\sigma \|g\|_{B^w_0(I_0)}.$$

If there holds $\|g\|_{B^w_0(I_0)} \lesssim \|g\|_w$, we are finished. If not, we take standard Calderón-Zygmund stopping data for g , \mathcal{G} and $\alpha_g(\cdot)$. By the corona estimate Theorem 6.6 in its dual form, we then have

$$|B^{\text{below}}(f, g)| \lesssim \mathcal{H} \|f\|_\sigma \|g\|_w + \sum_{G \in \mathcal{G}} |B^{\text{below}}_G(f, g)|$$

The estimate (6.15) applies to each term $B^{\text{below}}(Q^\sigma_G f, P^w_G g)$, for $G \in \mathcal{G}$. Quasi-orthogonality then completes the proof. \square

7. THE FUNCTIONAL ENERGY INEQUALITY

We state an important multi-scale extension of the energy inequality (5.3).

Definition 7.1. Let \mathcal{F} be a collection of dyadic intervals, and for each $F \in \mathcal{F}$. A collection of functions $\{g_F\}_{F \in \mathcal{F}}$ in $L^2(w)$ is said to be \mathcal{F}_ϵ -adapted if

- (1) The functions $g_F \in L^2_0(F, w)$.
- (2) Letting $\mathcal{J}(F) \equiv \{J : \widehat{g_F}(J) \neq 0\}$, these collections are pairwise disjoint in $F \in \mathcal{F}$.
- (3) For all $J \in \mathcal{J}(F)$ it holds that $J \Subset F$.
- (4) There is a finite number ρ so that for all intervals I the collection

$$(7.2) \quad \mathcal{K}(I) := \{J \subset I : J \text{ is maximal in } \mathcal{J}(F) \text{ for some } F \supset I\}$$

has bounded overlaps: $\left\| \sum_{J \in \mathcal{K}(I)} J(x) \right\|_\infty \leq \rho$.

Define \mathcal{F}_\subset similarly, with condition (3) replaced by

- (3') For all $J \in \mathcal{J}(F)$ it holds that $J \subset F$.

Concerning this definition, a natural choice for $\mathcal{J}(F)$ would be $\{J \Subset F : \pi_J F = F\}$, and one can check that this meets the definition above for $\rho = r$, the integer in the definition of $J \Subset F$. The more general definition will permit us to give a shorter proof of the parallel corona. We never need ρ more than a fixed constant, so we suppress the dependence of our estimates on this number.

Definition 7.3. Let \mathcal{F} be the smallest constant in the inequality below, or its dual form. The inequality holds for all non-negative $h \in L^2(\sigma)$, all σ -Carleson collections \mathcal{F} , and all \mathcal{F}_ϵ -adapted collections $\{g_F\}_{F \in \mathcal{F}}$:

$$\sum_{F \in \mathcal{F}} \sum_{J^* \in \mathcal{J}^*(F)} P(h\sigma, J^*) \left| \left\langle \frac{x}{|J^*|}, g_{F^*} \right\rangle_w \right| \leq \mathcal{F} \|h\|_\sigma \left[\sum_{F \in \mathcal{F}} \|g_F\|_w^2 \right]^{1/2}.$$

Here $\mathcal{J}^*(F)$ consists of the *maximal* intervals J in the collection $\mathcal{J}(F)$. Note that the estimate is universal in h and \mathcal{F} , separately.

This constant was identified in [9], and is herein shown to be necessary from the A_2 and interval testing inequalities.

Theorem 7.4. *There holds the inequality $\mathcal{F} \lesssim \mathcal{H}$.*

The first step in the proof is the domination of the constant \mathcal{F} by the best constant in a certain two weight inequality for the Poisson operator, with the weights being determined by w and σ in a particular way. This is the decisive step, since there is a two weight inequality for the Poisson operator proved by one of us. It reduces the full norm inequality to simpler testing conditions, which are in turn controlled by the A_2 and Hilbert transform testing conditions.

The way that the functional inequality is used is as in the following more technical corollary, which uses the weaker definition of \mathcal{F}_\subset -adapted.

Corollary 7.5. *Let $f \in L^2(\sigma)$ have Calderón-Zygmund stopping data \mathcal{F} and $\alpha_f(\cdot)$. For all \mathcal{F}_- -adapted functions $\{g_F\}$, there holds*

$$\left| \sum_{F \in \mathcal{F}} \sum_{I: I \supset F} \mathbb{E}_{I_F}^\sigma \Delta_I^\sigma f \cdot \langle H_\sigma I_F, g_F \rangle_w \right| \lesssim \mathcal{H} \|f\|_\sigma \left[\sum_{F \in \mathcal{F}} \|g_F\|_w^2 \right]^{1/2}.$$

The dual inequality also holds.

The proof of $\mathcal{F} \lesssim \mathcal{H}$ is taken up in the next few subsections, and then the proof of the corollary concludes this section.

7.1. The Two Weight Poisson Inequality. Consider the weight

$$\mu \equiv \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^*(F)} \left\| P_{F,J}^w \frac{x}{|J|} \right\|_w^2 \cdot \delta_{(x_J, |J|)}.$$

Here, $P_{F,J}^w := \sum_{J' \in \mathcal{J}(F): J' \subset J} \Delta_{J'}^w$. We can replace x by $x - c$ for any choice of c we wish; the projection is unchanged. And δ_q denotes a Dirac unit mass at a point q in the upper half plane \mathbb{R}_+^2 .

We prove the two-weight inequality for the Poisson integral:

$$\|\mathbb{P}(h\sigma)\|_{L^2(\mathbb{R}_+^2, \mu)} \lesssim \mathcal{H} \|h\|_\sigma,$$

for all nonnegative h . Above, $\mathbb{P}(\cdot)$ denotes the Poisson extension to the upper half-plane, so that in particular

$$\|\mathbb{P}(h\sigma)\|_{L^2(\mathbb{R}_+^2, \mu)}^2 = \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}(F)} \mathbb{P}(h\sigma)(x_J, |J|)^2 \left\| P_{F,J}^w \frac{x}{|J|} \right\|_w^2,$$

where x_J is the center of the interval J . The proof of Theorem 7.4 follows by duality.

Phrasing things in this way brings a significant advantage: The characterization of the two-weight inequality for the Poisson operator, [27], reduces the full norm inequality above to these testing inequalities. For any dyadic interval $I \in \mathcal{D}$

$$(7.6) \quad \int_{\mathbb{R}_+^2} \mathbb{P}(\sigma \cdot I)^2 d\mu(x, t) \lesssim \mathcal{H}^2 \sigma(I),$$

$$(7.7) \quad \int_{\mathbb{R}} \mathbb{P}^*(t\hat{I}\mu)^2 \sigma(dx) \lesssim \mathcal{A}_2 \int_{\hat{I}} t^2 \mu(dx, dt),$$

where $\hat{I} = I \times [0, |I|]$ is the box over I in the upper half-plane, and \mathbb{P}^* is the dual Poisson operator

$$\mathbb{P}^*(t\hat{I}\mu) = \int_{\hat{I}} \frac{t^2}{t^2 + |x - y|^2} \mu(dy, dt).$$

One should keep in mind that the intervals I are restricted to be in our fixed dyadic grid, a reduction allowed as the integrations on the left in (7.6) and (7.7) are done over the entire space, either \mathbb{R}_+^2 or \mathbb{R} . (Goodness of the intervals I above is not needed.) This reduction is critical to the analysis below.

Remark 7.8. A gap in the proof of the Poisson inequality at [27, Page 542] can be fixed as in [28] or [7].

7.2. The Poisson Testing Inequality: The Core. This subsection is concerned with this part of inequality (7.6): Restrict the integral on the left to the set $\hat{I} \subset \mathbb{R}_+^2$

$$\int_{\hat{I}} \mathbb{P}(\sigma \cdot I)^2 d\mu(x, t) \lesssim \mathcal{H}(\sigma(I)).$$

Since $(x_J, |J|) \in \hat{I}$ if and only if $J \subset I$, we have

$$\begin{aligned} \int_{\hat{I}} \mathbb{P}(\sigma \cdot I)(x, t)^2 d\mu(x, t) &= \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^*(F): J \subset I} \mathbb{P}(\sigma \cdot I)(x_J, |J|)^2 \left\| P_{F,J}^w \frac{x}{|J|} \right\|_w^2 \\ &\lesssim \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^*(F): J \subset I} \mathbb{P}(\sigma \cdot I, J)^2 \left\| P_{F,J}^w \frac{x}{|J|} \right\|_w^2. \end{aligned}$$

For each J ,

$$(7.9) \quad \left\| P_{F,J}^w \frac{x}{|J|} \right\|_w^2 \leq \int_J \left| \frac{x - \mathbb{E}_J^w x}{|J|} \right|^2 dw(x) = 2E(w, J)^2 w(J) \leq 2w(J).$$

A straight forward estimation is not possible, because the intervals J overlap. The intervals \mathcal{F} obey a σ -Carleson measure condition, which we exploit in the first stage of the proof. We 'create some holes' by restricting the support of σ to the interval I in the sum below.

$$\begin{aligned} &\sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^*(F): J \subset I} \mathbb{P}((F \cap I)\sigma, J)^2 \left\| P_{F,J}^w \frac{x}{|J|} \right\|_w^2 \\ &= \left\{ \sum_{F \in \mathcal{F}: F \subset I} + \sum_{F \in \mathcal{F}: F \not\subset I} \right\} \sum_{J \in \mathcal{J}^*(F): J \subset I} \mathbb{P}((F \cap I)\sigma, J)^2 \left\| P_{F,J}^w \frac{x}{|J|} \right\|_w^2 \\ &= A + B. \end{aligned}$$

The first of these terms is at most

$$\begin{aligned} A &\leq \sum_{F \in \mathcal{F}: F \subset I} \sum_{J \in \mathcal{J}^*(F)} \mathbb{P}(F\sigma, J)^2 E(w, J)^2 w(J) \\ &\leq \mathcal{H}^2 \sum_{F \in \mathcal{F}: F \subset I} \sigma(F) \lesssim \mathcal{H}^2 \sigma(I). \end{aligned}$$

Here we have used (7.9), the energy inequality (5.3), and that the stopping intervals \mathcal{F} satisfy a σ -Carleson measure estimate property (2) of Definition 3.4.

Concerning the second term, this is the point that the element of the definition (7.2) enters into the proof. We can write the collection $\mathcal{K}(I)$ of (7.2) as the union of collections \mathcal{K}_k , $1 \leq k \leq \rho$, where the intervals in each \mathcal{K}_k are a subpartition of I . Using (7.9) and the energy inequality, term

B satisfies

$$\begin{aligned} B &\leq \sum_{k=1}^{\rho} \sum_{J \in \mathcal{K}_k} P(\sigma \cdot I, J)^2 \left\| P_{F_J}^w \frac{\chi}{|J|} \right\|_w^2 \\ &\lesssim \sum_{k=1}^{\rho} \sum_{J \in \mathcal{K}_k} P(\sigma \cdot I, J)^2 E(w, J)^2 w(J) \lesssim \rho \mathcal{H}^2 \sigma(I). \end{aligned}$$

In the first line, F_J is the unique $F \in \mathcal{F}$ with $J \in \mathcal{J}^*(F)$ and $F \supset I$.

It remains then to show the following inequality with ‘holes’:

$$\sum_{F \in \mathcal{F}_I} \sum_{J \in \mathcal{J}^*(F)} P(\sigma(I - F), J)^2 \left\| P_{F_J}^w \frac{\chi}{|J|} \right\|_w^2 \lesssim \mathcal{H}^2 \sigma(I),$$

where \mathcal{F}_I consists of those $F \in \mathcal{F}$ with $F \subset I$. Our purpose is to pass back to the Hilbert transform, so that we can effectively use the testing condition. The inequality above can be expressed in dual language as the inequality

$$\sum_{F \in \mathcal{F}_I} \sum_{J \in \mathcal{J}^*(F)} P(\sigma(I - F), J) \left\langle P_{F_J}^w \frac{\chi}{|J|}, g \right\rangle_w \lesssim \mathcal{H} \sigma(I)^{1/2} \|g\|_w.$$

In the inner product, $g \in L^2(w)$ can be replaced by $g_{F_J} := P_{F_J}^w g$ by self-adjointness of the projections. Also, $\langle \chi, h_J^w \rangle_w > 0$, so that we are free to assume that $\langle g, h_J^w \rangle_w \geq 0$ for all J .

We can estimate, using the monotonicity property (5.2),

$$P(\sigma(I - F), J) \left\langle \frac{\chi}{|J|}, g_{F_J} \right\rangle_w \approx \langle H_\sigma(I - F), g_{F_J} \rangle_w, \quad J \in \mathcal{J}(F).$$

It therefore suffices to show that

$$(7.10) \quad \sum_{F \in \mathcal{F}_I} \sum_{J \in \mathcal{J}(F)} \langle H_\sigma(I - F), g_{F_J} \rangle_w \lesssim \mathcal{H} \sigma(I)^{1/2} \|g\|_w.$$

Use linearity in the argument of the Hilbert transform, which gives two terms. The first is

$$\left| \sum_{F \in \mathcal{F}_I} \sum_{J \in \mathcal{J}(F)} \langle H_\sigma I, g_{F_J} \rangle_w \right| = |\langle H_\sigma I, g \rangle_w| \leq \mathcal{H} \sigma(I)^{1/2} \|g\|_w.$$

The second term appeals to interval testing and the σ -Carleson measure condition (3.3).

$$\begin{aligned} \left| \sum_{F \in \mathcal{F}_I} \sum_{J \in \mathcal{J}(F)} \langle H_\sigma F, g_{F_J} \rangle_w \right| &\leq \sum_{F \in \mathcal{F}_I} \left| \left\langle H_\sigma F, \sum_{J \in \mathcal{J}(F)} g_{F_J} \right\rangle_w \right| \\ &\leq \mathcal{H} \sum_{F \in \mathcal{F}_I} \sigma(F)^{1/2} \left\| \sum_{J \in \mathcal{J}(F)} g_{F_J} \right\|_w \\ &\leq \mathcal{H} \left[\sum_{F \in \mathcal{F}_I} \sigma(F) \times \sum_{F \in \mathcal{F}_I} \sum_{J \in \mathcal{J}(F)} \|g_{F_J}\|_w^2 \right]^{1/2} \end{aligned}$$

$$\lesssim \mathcal{H}\sigma(I)^{1/2} \|g\|_w.$$

We also use the orthogonality of the functions $g_{F,J}$. This completes the proof of (7.10).

All the remaining estimates use the orthogonality of $g_{F,J}$ and the \mathcal{A}_2 condition. The details are below.

7.3. The Poisson Testing Inequality: The Remainder. Now we turn to proving the following estimate for the global part of the first testing condition (7.6):

$$\int_{\mathbb{R}_+^2 - \widehat{I}} \mathbb{P}(\sigma \cdot I)^2 d\mu \lesssim \mathcal{A}_2 \sigma(I).$$

Decomposing the integral on the left into four terms: With F_J the unique $F \in \mathcal{F}$ with $J \in \mathcal{J}^*(F)$,

$$\begin{aligned} \int_{\mathbb{R}_+^2 - \widehat{I}} \mathbb{P}(\sigma \cdot I)^2 d\mu &= \sum_{J: (x_J, |J|) \in \mathbb{R}_+^2 - \widehat{I}} \mathbb{P}(\sigma \cdot I)(x_J, |J|)^2 \left\| P_{F_J, J}^w \frac{x}{|J|} \right\|_w^2 \\ &= \left\{ \sum_{\substack{J: J \cap 3I = \emptyset \\ |J| \leq |I|}} + \sum_{J: J \subset 3I - I} + \sum_{\substack{J: J \cap I = \emptyset \\ |J| > |I|}} + \sum_{J: J \not\supset I} \right\} \mathbb{P}(\sigma \cdot I)(x_J, |J|)^2 \left\| P_{F_J, J}^w \frac{x}{|J|} \right\|_w^2 \\ &= A + B + C + D. \end{aligned}$$

Decompose term A according to the length of J and its distance from I , and then use (7.9) to obtain:

$$\begin{aligned} A &\lesssim \sum_{n=0}^{\infty} \sum_{k=1}^{\infty} \sum_{\substack{J: J \subset 3^{k+1}I - 3^kI \\ |J| = 2^{-n}|I|}} \left(\frac{2^{-n}|I|}{\text{dist}(J, I)^2} \sigma(I) \right)^2 w(J) \\ &\lesssim \sum_{n=0}^{\infty} 2^{-2n} \sum_{k=1}^{\infty} \frac{|I|^2 \sigma(I) w(3^{k+1}I - 3^kI)}{|3^kI|^4} \sigma(I) \\ &\lesssim \sum_{n=0}^{\infty} 2^{-2n} \sum_{k=1}^{\infty} 3^{-2k} \left\{ \frac{\sigma(3^{k+1}I) w(3^{k+1}I)}{|3^kI|^2} \right\} \sigma(I) \lesssim \mathcal{A}_2 \sigma(I). \end{aligned}$$

Decompose term B according to the length of J and then use the Poisson inequality (5.5), available to use because of goodness of intervals J . We then obtain

$$\begin{aligned} B &\lesssim \sum_{n=0}^{\infty} \sum_{\substack{J: J \subset 3I - I \\ |J| = 2^{-n}|I|}} 2^{-n(2-4\epsilon)} \left(\frac{\sigma(I)}{|I|} \right)^2 w(J) \\ &\leq \sum_{n=0}^{\infty} 2^{-n(2-4\epsilon)} \frac{\sigma(3I) w(3I)}{|3I|} \sigma(I) \lesssim \mathcal{A}_2 \sigma(I). \end{aligned}$$

For term C , split the sum according to whether or not I intersects the triple of J :

$$\begin{aligned} C &\lesssim \left\{ \sum_{\substack{J: I \cap 3J = \emptyset \\ |J| > |I|}} + \sum_{\substack{J: I \subset 3J - J \\ |J| > |I|}} \right\} \left(\frac{|J| \sigma(I)}{\text{dist}(J, I)^2} \right)^2 \left\| P_{F_J, J}^w \frac{x}{|J|} \right\|_w^2 \\ &= C_1 + C_2. \end{aligned}$$

To estimate C_1 , let $\{B_i\}_{i=1}^\infty$ be the maximal intervals in the collection of triples

$$\{3J : |J| > |I| \text{ and } 3J \cap I = \emptyset\},$$

arranged in order of increasing side length. These intervals are not disjoint, but have bounded overlap, $\sum_{i=1}^\infty B_i \leq 3$. Group the intervals J by the inclusion $3J \subset B_i$, appeal to $P_{F_J}^w x = P_{F_J}^w (x - c(B_i))$, and use the mutual orthogonality of the $P_{F_J}^w$:

$$\begin{aligned} C_1 &\leq \sum_{i=1}^\infty \sum_{J: 3J \subset B_i} \left(\frac{|J| \sigma(I)}{\text{dist}(J, I)^2} \right)^2 \left\| P_{F_J, J}^w \frac{x}{|J|} \right\|_w^2 \\ &\lesssim \sum_{i=1}^\infty \left(\frac{\sigma(I)}{\text{dist}(B_i, I)^2} \right)^2 \sum_{J: 3J \subset B_i} \left\| P_{F_J, J}^w x \right\|_w^2 \\ &\lesssim \sum_{i=1}^\infty \left(\frac{\sigma(I)}{\text{dist}(B_i, I)^2} \right)^2 \|B_i (x - c(B_i))\|_w^2 \\ &\lesssim \sum_{i=1}^\infty \left(\frac{\sigma(I)}{\text{dist}(B_i, I)^2} \right)^2 |B_i|^2 w(B_i) \\ &\lesssim \left\{ \sum_{i=1}^\infty \frac{w(B_i) \sigma(I)}{|B_i|^2} \right\} \sigma(I) \lesssim \mathcal{A}_2 \sigma(I). \end{aligned}$$

Next we turn to estimating term C_2 where the triple of J contains I but J itself does not. Note that there are at most two such intervals J of a given length, one to the left and one to the right of I . So with this in mind we sum over the intervals J according to their lengths and use (7.9) to obtain

$$\begin{aligned} C_2 &= \sum_{n=0}^\infty \sum_{\substack{J: I \subset 3J - J \\ |J| = 2^n |I|}} \left(\frac{|J| \sigma(I)}{\text{dist}(J, I)^2} \right)^2 \left\| P_{F_J, J}^w \frac{x}{|J|} \right\|_w^2 \\ &\lesssim \sum_{n=0}^\infty \left(\frac{\sigma(I)}{|2^n I|} \right)^2 w(3 \cdot 2^n I) = \left\{ \frac{\sigma(I)}{|I|} \sum_{n=0}^\infty \frac{w(3 \cdot 2^n I)}{|2^n I|^2} \right\} \sigma(I) \\ &\lesssim \left\{ \frac{\sigma(I)}{|I|} P(w, I) \right\} \sigma(I) \leq \mathcal{A}_2 \sigma(I). \end{aligned}$$

The last term D is handled in the same way as term C_2 . The intervals J occurring here are included in the set of ancestors $A_k \equiv \pi_D^k I$ of I , $1 \leq k < \infty$. We thus have

$$\begin{aligned} D &= \sum_{k=1}^{\infty} \mathbb{P}(\sigma \cdot I) (c(A_k), |A_k|)^2 \left\| P_{F_{A_k}, A_k}^w \frac{x}{|A_k|} \right\|_w^2 \\ &\lesssim \sum_{k=1}^{\infty} \left(\frac{\sigma(I)}{|A_k|} \right)^2 w(A_k) = \left\{ \frac{\sigma(I)}{|I|} \sum_{k=1}^{\infty} \frac{|I|}{|A_k|^2} w(A_k) \right\} \sigma(I) \\ &\lesssim \left\{ \frac{\sigma(I)}{|I|} P(w, I) \right\} \sigma(I) \lesssim \mathcal{A}_2 \sigma(I). \end{aligned}$$

7.4. The Dual Poisson Testing Inequality. We are considering (7.7). Note that the expressions on the two sides of this inequality are

$$\begin{aligned} \int_{\hat{I}} t^2 \mu(dx, dt) &= \sum_{F \in \mathcal{F}} \sum_{\substack{J \in \mathcal{J}^*(F) \\ J \subset I}} \|P_{F,J}^w x\|_w^2, \\ \mathbb{P}^*(t\hat{I}\mu)(x) &= \sum_{F \in \mathcal{F}} \sum_{\substack{J \in \mathcal{J}^*(F) \\ J \subset I}} \frac{\|P_{F,J}^w x\|_w^2}{|J|^2 + |x - x_J|^2}. \end{aligned}$$

Some bookkeeping is in order. We take $J \in \mathcal{J}^e(F)$ iff $J \subset I$, and F is the maximal interval in \mathcal{F} with $J \in \mathcal{J}^*(F)$. Then each interval J is in at most one collection $\mathcal{J}^e(F)$. Below, we understand that sum over the empty set to be the zero projection. Define

$$Q_J^w = \sum_{F \in \mathcal{F} : J \in \mathcal{J}^*(F)} P_{F,J}^w.$$

These are mutually orthogonal projections, and are indexed solely by $J \in \mathcal{D}$. By the mutual orthogonality of the $P_{F,J}^w$, the quantities above are equal to

$$\begin{aligned} \int_{\hat{I}} t^2 \mu(dx, dt) &= \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^e(F)} \|Q_J^w x\|_w^2, \\ \mathbb{P}^*(t\hat{I}\mu)(x) &= \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^e(F)} \frac{\|Q_J^w x\|_w^2}{|J|^2 + |x - x_J|^2}. \end{aligned}$$

We are to dominate $\|\mathbb{P}^*(t\hat{I}\mu)\|_w^2$ by the first expression above. Expanding the squared norm, the diagonal term is

$$\begin{aligned} \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^e(F)} \int \left[\frac{\|Q_J^w x\|_w^2}{|J|^2 + |x - x_J|^2} \right]^2 d\sigma &\leq M_1 \cdot \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^e(F)} \|Q_J^w x\|_w^2 \\ \text{where } M_1 &\equiv \sup_{F \in \mathcal{F}} \sup_{J \in \mathcal{J}^e(F)} \int \frac{\|Q_J^w x\|_w^2}{(|J|^2 + |x - x_J|^2)^2} d\sigma. \end{aligned}$$

But, by inspection, M_1 is dominated by the \mathcal{A}_2 constant. Indeed, for any J , we have by (7.9)

$$\int \frac{\|Q_J^w x\|_w^2}{(|J|^2 + |x - x_J|^2)^2} d\sigma \leq \frac{w(J)}{|J|} \int \frac{|J|^3}{(|J|^2 + |x - x_J|^2)^2} \sigma(dx) \leq \mathcal{A}_2.$$

Having fixed ideas, we fix an integer s , and consider those intervals $J, J' \in \mathcal{J}^\epsilon(F)$ with $|J'| = 2^{-s}|J|$. The expression to control is

$$\begin{aligned} T_s &:= \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^\epsilon(F)} \sum_{\substack{F' \in \mathcal{F} \\ F' \neq F}} \sum_{\substack{J' \in \mathcal{J}^\epsilon(F) \\ |J'| = 2^{-s}|J|}} \int_I \frac{\|Q_J^w x\|_w^2}{|J|^2 + |x - x_J|^2} \cdot \frac{\|Q_{J'}^w x\|_w^2}{|J'|^2 + |x - c(J')|^2} d\sigma \\ &\leq M_2 \sum_{F \in \mathcal{F}} \sum_{J \in \mathcal{J}^\epsilon(F)} \|Q_J^w x\|_w^2 \end{aligned}$$

$$\text{where } M_2 \equiv \sup_{F \in \mathcal{F}} \sup_{J \in \mathcal{J}^\epsilon(F)} \sum_{\substack{F' \in \mathcal{F} \\ F' \neq F}} \sum_{\substack{J' \in \mathcal{J}^\epsilon(F) \\ |J'| = 2^{-s}|J|}} \int_I \frac{1}{|J|^2 + |x - x_J|^2} \cdot \frac{\|Q_{J'}^w x\|_w^2}{|J'|^2 + |x - c(J')|^2} d\sigma.$$

We claim the term M_2 is at most a constant times $\mathcal{A}_2 2^{-s}$. To see, fix J as in the definition of M_2 , and use (7.9) to estimate the integral on the right by

$$\frac{w(J')}{|J'|} \int_I \frac{|J'|^2}{|J|^2 + |x - x_J|^2} \cdot \frac{|J'|}{|J'|^2 + |x - c(J')|^2} d\sigma \lesssim \mathcal{A}_2 \frac{2^{-2s}}{1 + n^2}$$

where n is an integer chosen so that $(n-1)|J| \leq \text{dist}(J, J') \leq n|J|$. Then estimate the sum over J' as follows.

$$\sum_{F' \in \mathcal{F}} \sum_{\substack{J' \in \mathcal{J}^\epsilon(F') : |J'| = 2^{-s}|J| \\ (n-1)|J| \leq \text{dist}(J, J') \leq n|J|}} \frac{2^{-2s}}{1 + n^2} \lesssim \frac{2^{-s}}{1 + n^2}.$$

because the relative lengths of J and J' are fixed, and each J' is in at most one $\mathcal{J}^\epsilon(F)$. This is summable over $n \in \mathbb{N}$ to 2^{-s} , so it completes our proof.

7.5. Proof of Corollary 7.5. Write $g_F = g_F^1 + g_F^2$, where $g_F^1 := \sum_{J \in \mathcal{J}(F) : J \notin F} \Delta_J^w g$. We argue the case of g_F^2 first. The functions $\{g_F^2 : F \in \mathcal{F}\}$ are \mathcal{F}_ϵ -adapted, with the only point not being immediate is the technical condition (7.2). Take interval I , and pair (J, F) with $J \subset I \subset F$, and J maximal in $\mathcal{J}(F)$. It must be that $\pi_{\mathcal{F}} J = F$, hence $F \subset \pi_{\mathcal{F}}^r I$. That is, F can only take at most r possible values in \mathcal{F} . Hence, this condition holds with $\rho = r$.

This argument only depends upon the functional energy inequality. The argument of the Hilbert transform is I_F , the child of I that contains F . Write $I_F = F + (I_F - F)$, and use linearity of H_σ . Note that by the standard martingale difference identity and the construction of stopping data,

$$\left| \sum_{I : I \sqsubset F} \mathbb{E}_{I_F}^\sigma \Delta_I^\sigma f \right| \lesssim \alpha_f(F), \quad F \in \mathcal{F}.$$

Hence, the first term is

$$\begin{aligned} \left| \sum_{F \in \mathcal{F}} \sum_{I: I \supset F} \mathbb{E}_{I_F}^\sigma \Delta_I^\sigma f \cdot \langle H_\sigma F, g_F^2 \rangle_w \right| &\lesssim \sum_{F \in \mathcal{F}} \alpha_f(F) \left| \langle H_\sigma F, g_F^2 \rangle_w \right| \\ &\lesssim \mathcal{H} \sum_{F \in \mathcal{F}} \alpha_f(F) \sigma(F)^{1/2} \|g_F^2\|_w. \end{aligned}$$

This just uses interval testing. Quasi-orthogonality bounds this last expression.

For the second expression, when the argument of the Hilbert transform is $I_F - F$, first note that

$$\left| \sum_{I: I \supset F} \mathbb{E}_{I_F}^\sigma \Delta_I^\sigma f \cdot (I_F - F) \right| \lesssim \Phi := \sum_{F' \in \mathcal{F}} \alpha_f(F') \cdot F', \quad F \in \mathcal{F}.$$

Therefore, by the definition of \mathcal{F} adapted, the monotonicity property (5.2) applies, and yields

$$\left| \sum_{I: I \supset F} \mathbb{E}_{I_F}^\sigma \Delta_I^\sigma f \cdot \langle H_\sigma(I_F - F), g_F^2 \rangle_w \right| \lesssim \sum_{J \in \mathcal{J}^*(F)} P(\Phi \sigma, J) \left\langle \frac{x}{|J|}, J g_F^2 \right\rangle_w, \quad F \in \mathcal{F}.$$

The sum over $F \in \mathcal{F}$ of this last expression is controlled by functional energy, and the property that $\|\Phi\|_\sigma \lesssim \|f\|_\sigma$.

Remark 7.11. The proof above is the paraproduct trick of Nazarov-Treil-Volberg. Applying this trick to stopping intervals for f eliminates the need to verify that certain derived measures are σ -Carleson measures, a delicate part of the arguments in [8, 17, 29].

We return to the functions g_F^1 defined at the beginning of the proof; the Haar supports of these functions are ‘close to F ’. Define functions

$$\tilde{g}_F^s := \sum_{F': \pi_{\mathcal{F}}^s F' = F} g_{F'}^1, \quad s \in \mathbb{N}.$$

Here, the sum is over F' which are s steps below F in the \mathcal{F} tree. It is straight forward to verify that the functions $\{\tilde{g}_F^{r+1}\}$ are also \mathcal{F}_ϵ -adapted. Hence, by the argument for g_F^2 , there holds

$$\left| \sum_{F \in \mathcal{F}} \sum_{I: I \supset F} \mathbb{E}_{I_F}^\sigma \Delta_I^\sigma f \cdot \langle H_\sigma I_F, \tilde{g}_F^{r+1} \rangle_w \right| \lesssim \mathcal{H} \|f\|_\sigma \left[\sum_{F \in \mathcal{F}} \|g_F^2\|_w^2 \right]^{1/2}.$$

It remains to establish the estimate below, uniform over $1 \leq s \leq r$ and $F \in \mathcal{F}$.

$$\left| \sum_{I: F \subsetneq I \subset \pi_{\mathcal{F}} F} \mathbb{E}_{I_F}^\sigma \Delta_I^\sigma f \cdot \langle H_\sigma I_F, \tilde{g}_F^s \rangle_w \right| \lesssim \mathcal{H} \alpha_f(F) \sigma(F)^{1/2} \|\tilde{g}_F^s\|_w^2.$$

For then, quasi-orthogonality completes the case. The distinction here is that we need not use functional energy.

There is an elementary subcase. In [8, Proposition 2.8], it is shown that for any $A \geq 1$,

$$\sup_{\substack{I, J: I \cap J \neq \emptyset \\ |J| \leq A|I| \leq A^2|J|}} |\langle H_\sigma I, J \rangle_w| \leq C_A \mathcal{H} \sqrt{\sigma(I)w(I)}.$$

Apply this with $A = 2^{2r}$ to see that the estimate above holds for the sum below, where the relative lengths of I and J are controlled.

$$\sum_{I: F \subseteq I \subset \pi_{\mathcal{F}} F} \sum_{\substack{J: J \subseteq F \\ |I| \leq 2^{2r}|J|}} \mathbb{E}_{I_F}^\sigma \Delta_I^\sigma f \cdot \langle H_\sigma I_F, \Delta_J^w \tilde{g}_F^s \rangle_w.$$

It remains to consider the complementary sum. In this case, we return to the paraproduct trick mentioned above. Let $\tilde{\mathcal{F}}$ be the intervals $F' \in \mathcal{F}$ with $\pi_{\mathcal{F}}^s F' = F$. Then, for each $F' \in \tilde{\mathcal{F}}$, we write the argument of the Hilbert transform as $I_F = F + (I_F - F)$. Interval testing shows that

$$\sum_{I: F \subseteq I \subset \pi_{\mathcal{F}} F} \sum_{\substack{J: J \subseteq F \\ |I| \geq 2^{2r}|J|}} \mathbb{E}_{I_F}^\sigma \Delta_I^\sigma f \cdot \langle H_\sigma F, \Delta_J^w \tilde{g}_F^s \rangle_w$$

satisfies the correct bound. For the second term, use the monotonicity property to see that

$$\left| \sum_{I: F \subseteq I \subset \pi_{\mathcal{F}} F} \sum_{\substack{J: J \subseteq F \\ |I| \geq 2^{2r}|J|}} \mathbb{E}_{I_F}^\sigma \Delta_I^\sigma f \cdot \langle H_\sigma (I_F - F), \Delta_J^w \tilde{g}_F^s \rangle_w \right| \lesssim \sum_{F' \in \tilde{\mathcal{F}}} \sum_{J \in \mathcal{J}(F')} P(\sigma F, J) E(w, J) \|P_J^w g_F^s\|_w,$$

where $\mathcal{J}(F')$ is the maximal intervals J in the Haar support of $g_{F'}$ so that there is an $I \supset F$ with $2^{2r}|J| < |I|$. Then, $P_J^w = \sum_{J': J' \subset J} \Delta_{J'}^w$. Cauchy-Schwarz and the energy inequality (5.3) concludes this estimate.

REFERENCES

- [1] M. Cotlar and C. Sadosky, *On some L^p versions of the Helson-Szegő theorem*, (Chicago, Ill., 1981), Wadsworth Math. Ser., Wadsworth, Belmont, CA, 1983, pp. 306–317. MR730075 (85i:42015)
- [2] Guy David and Jean-Lin Journé, *A boundedness criterion for generalized Calderón-Zygmund operators*, Ann. of Math. (2) **120** (1984), no. 2, 371–397. MR763911 (85k:42041)
- [3] V. P. Havin and N. K. Nikolski (eds.), *Linear and complex analysis. Problem book 3. Part I*, Lecture Notes in Mathematics, vol. 1573, Springer-Verlag, Berlin, 1994. MR1334345 (96c:00001a)
- [4] Tuomas P. Hytönen, *The sharp weighted bound for general Calderón-Zygmund operators*, Ann. of Math. (2) **175** (2012), no. 3, 1473–1506.
- [5] Tuomas P. Hytönen, Michael T. Lacey, Henri Martikainen, Tuomas Orponen, Maria Carmen Reguera, Eric T. Sawyer, and Ignacio Uriarte-Tuero, *Weak and strong type estimates for maximal truncations of Calderón-Zygmund operators on A_p weighted spaces*, J. Anal. Math. **118** (2012), 177–220.
- [6] Michael T. Lacey, Stefanie Petermichl, and Maria Carmen Reguera, *Sharp A_2 inequality for Haar shift operators*, Math. Ann. **348** (2010), no. 1, 127–141. MR2657437
- [7] Michael T. Lacey, Eric T. Sawyer, and Ignacio Uriarte-Tuero, *A characterization of two weight norm inequalities for maximal singular integrals with one doubling measure*, Accepted to Analysis and PDE (2008), available at <http://www.arxiv.org/abs/0807.0246>.
- [8] ———, *A Two Weight Inequality for the Hilbert transform Assuming an Energy Hypothesis*, J Funct Anal **263** (2012), 305–363.
- [9] Michael T. Lacey, Eric T. Sawyer, Chun-Yun Shen, and Ignacio Uriarte-Tuero, *The two weight inequality for the Hilbert transform, coronas, and energy conditions* (2011), available at <http://www.arXiv.org/abs/1108.2319>.
- [10] Yurii I. Lyubarskii and Kristian Seip, *Complete interpolating sequences for Paley-Wiener spaces and Muckenhoupt's (A_p) condition*, Rev. Mat. Iberoamericana **13** (1997), no. 2, 361–376. MR1617649 (99e:42004)

- [11] Benjamin Muckenhoupt, *Weighted norm inequalities for the Hardy maximal function*, Trans. Amer. Math. Soc. **165** (1972), 207–226. MR0293384 (45 #2461)
- [12] ———, *Hardy's inequality with weights*, Studia Math. **44** (1972), 31–38. Collection of articles honoring the completion by Antoni Zygmund of 50 years of scientific activity, I. MR0311856 (47 #418)
- [13] F. Nazarov, *A counterexample to Sarason's conjecture*, Preprint, MSU (1997), available at <http://www.math.msu.edu/~fedja/prepr.html>.
- [14] F. Nazarov, S. Treil, and A. Volberg, *Cauchy integral and Calderón-Zygmund operators on nonhomogeneous spaces*, Internat. Math. Res. Notices **15** (1997), 703–726.
- [15] ———, *Accretive system Tb-theorems on nonhomogeneous spaces*, Duke Math. J. **113** (2002), no. 2, 259–312.
- [16] ———, *The Tb-theorem on non-homogeneous spaces*, Acta Math. **190** (2003), no. 2, 151–239.
- [17] ———, *Two weight estimate for the Hilbert transform and Corona decomposition for non-doubling measures* (2004), available at <http://arxiv.org/abs/1003.1596>.
- [18] ———, *Two weight inequalities for individual Haar multipliers and other well localized operators*, Math. Res. Lett. **15** (2008), no. 3, 583–597. MR2407233 (2009e:42031)
- [19] F. Nazarov and A. Volberg, *The Bellman function, the two-weight Hilbert transform, and embeddings of the model spaces K_θ* , J. Anal. Math. **87** (2002), 385–414. Dedicated to the memory of Thomas H. Wolff.
- [20] Nikolai Nikolski and S. Treil, *Linear resolvent growth of rank one perturbation of a unitary operator does not imply its similarity to a normal operator*, J. Anal. Math. **87** (2002), 415–431. Dedicated to the memory of Thomas H. Wolff. MR1945291 (2004e:47004)
- [21] Carlos Pérez, S. Treil, and A. Volberg, *On A_2 conjecture and Corona decomposition of weights* (2010), available at <http://arxiv.org/abs/1006.2630>.
- [22] ———, *Sharp weighted estimates for dyadic shifts and the A_2 conjecture* (2010), available at <http://arxiv.org/abs/1010.0755>.
- [23] Alexei Poltoratski and Donald Sarason, *Aleksandrov-Clark measures*, Recent advances in operator-related function theory, Contemp. Math., vol. 393, Amer. Math. Soc., Providence, RI, 2006, pp. 1–14. MR2198367 (2006i:30048)
- [24] Maria Carmen Reguera and James Scurry, *On joint estimates for maximal functions and singular integrals in weighted spaces*, PAMS, to appear (2011), available at <http://www.arxiv.org/abs/1109.2027>.
- [25] Donald Sarason, *Exposed points in H^1 . II*, Topics in operator theory: Ernst D. Hellinger memorial volume, Oper. Theory Adv. Appl., vol. 48, Birkhäuser, Basel, 1990, pp. 333–347. MR1207406 (94a:46031)
- [26] Eric T. Sawyer, *A characterization of a two-weight norm inequality for maximal operators*, Studia Math. **75** (1982), no. 1, 1–11. MR676801 (84i:42032)
- [27] ———, *A characterization of two weight norm inequalities for fractional and Poisson integrals*, Trans. Amer. Math. Soc. **308** (1988), no. 2, 533–545. MR930072 (89d:26009)
- [28] E. Sawyer and R. L. Wheeden, *Weighted inequalities for fractional integrals on Euclidean and homogeneous spaces*, Amer. J. Math. **114** (1992), no. 4, 813–874. MR1175693 (94i:42024)
- [29] A. Volberg, *Calderón-Zygmund capacities and operators on nonhomogeneous spaces*, CBMS Regional Conference Series in Mathematics, vol. 100, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 2003.

SCHOOL OF MATHEMATICS, GEORGIA INSTITUTE OF TECHNOLOGY, ATLANTA GA 30332

E-mail address: lacey@math.gatech.edu

DEPARTMENT OF MATHEMATICS & STATISTICS, MCMASTER UNIVERSITY, 1280 MAIN STREET WEST,
HAMILTON, ONTARIO, CANADA L8S 4K1

E-mail address: sawyer@mcmaster.ca

DEPARTMENT OF MATHEMATICS & STATISTICS, MCMASTER UNIVERSITY, 1280 MAIN STREET WEST,
HAMILTON, ONTARIO, CANADA L8S 4K1

E-mail address: cyshen@math.mcmaster.ca

DEPARTMENT OF MATHEMATICS, MICHIGAN STATE UNIVERSITY, EAST LANSING MI

E-mail address: ignacio@math.msu.edu